

University of St Andrews



Full metadata for this thesis is available in
St Andrews Research Repository
at:

<http://research-repository.st-andrews.ac.uk/>

This thesis is protected by original copyright

Moral Minimalism

by

Lars Grassmé Binderup

University of St. Andrews

Ph.D. thesis

Submitted 8th of September 2000



Declarations

I, Lars Grassmé Binderup, hereby certify that this thesis, which is approximately 98.000 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,

date. 2/1/01...signature of candidate..

I was admitted as a research student in January 1996 and as a candidate for the degree of Ph.D. in Logic and Metaphysics; the higher study for which this is a record was carried out in the University of St. Andrews between 1996 and 2000.

date. 2/1/01...signature of candidate...

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

date. 12/1/2001/signature of supervisor..(

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.

date. 2/1/01...signature of candidate.

Synopsis

The thesis is a defence of a novel – minimalist – position in meta-ethics. The minimalist claims that we can defend the truth-aptness of moral discourse without becoming objectivists and or realists about morals. Thus, the position can be seen as an attempt to formulate a consistent cognitivist version of moral relativism. The thesis is formulated using Crispin Wright's framework for realism-debate from his *Truth and Objectivity* (Harvard University Press, 1992).

Minimalism is, first of all, the claim that moral discourse is *minimally truth-apt*, since moral discourse displays the appropriate syntax and discipline – where the account of the latter involves the claim that warrant for moral beliefs consists in affectively grounded, defeasible, entitlements.

However, the pluralist account of truth advocated in the thesis makes it possible to claim that truth in morals has no *further* realism or objectivity-relevant features. Thus, it is argued that moral truth is epistemically constrained. Also, the affective grounding of moral discourse – and the ensuing implausibility of an intuitional moral epistemology – means that we cannot have an a priori guarantee that any moral disagreement involves a cognitive shortcoming on the part of one or both disputants. This amounts to the claim that moral discourse lacks *Cognitive Command*.

Further it is argued that no *substantively naturalistic account of morals* – whether analytical or ontological – nor any argument from the phenomenon of *moral-on-natural substantive supervenience* has so far succeeded in showing that moral discourse is linked to naturalistic discourse in a way that can ensure the objectivity of morals. That is, moral discourse does not *disputationally supervene* on naturalistic discourse.

Minimalism is, it is argued, consistent with phenomena like *radical moral disagreement* and *moral dilemmas*. It is compatible with the normative and motivational character of moral beliefs given that it is possible to defend a notion of *minimal belief*. Finally, moral minimalism is compatible with a *methodologically* naturalistic account of moral discourse.

Table of Contents

Acknowledgements	1
Introduction	2
Chapter One: The Stability of Moral Minimalism	7
1.1 Moral minimalism and deflationism.	8
1.2 Minimal truth and the debates about moral objectivity and realism.	12
1.3 Minimal truth-aptitude, discipline and moral discourse.	16
1.4 Superassertibility as truth in morals.	22
1.5 An argument for Epistemic Constraint on moral truth.	26
1.6 Moral minimalism summarised.	31
1.7 Introducing the Cognitive Command constraint.	32
1.8 Cognitive Command and Epistemic Constraint.	37
1.9 Trivialising the Cognitive Command constraint?	39
1.10 Trivialising Cognitive Command: Two arguments.	43
1.11 Meeting the trivialising challenges: Agnosticism about classical logic.	47
1.12 Meeting the trivialising challenges: Characterising indeterminacy.	50
1.13 Conceptual knowledge and modality: A final objection?	53
Chapter Two: Minimalism, Expressivism and Motivation	58
2.1 Minimalism and expressivism: The argument for compatibility.	60
2.2 Humeanism about motivating and normative reasons.	63
2.3 Three ways of resisting the standard argument for expressivism.	67
2.4 Minimalism and minimal belief.	77
2.5 Answering the Australian challenge.	82

Chapter Three: Moral Minimalism and Moral Disagreement **85**

3.1	What is an un-settleable dispute?	87
3.2	De facto versus absolute un-settleability.	92
3.3	Examples of radical moral disputes.	94
3.4	The univocality of radical moral disagreement and the core platitudes.	103
3.5	First proposal for characterising actual moral disagreement – Scylla.	108
3.6	Second proposal for characterising actual moral disagreement – Charybdis.	113
3.7	Minimalism and moral disagreement.	118

Chapter Four: Minimalism and Moral Dilemmas **121**

4.1	Two challenges.	122
4.2	Moral conflict and moral disagreement.	124
4.3	Moral conflict and inconsistency.	126
4.4	Towards a characterisation of moral dilemmas.	128
4.5	Dilemmas, deontology and teleology.	130
4.6	Tragic moral dilemmas, moral residue and un-settleability.	134
4.7	Matched moral reasons and a revised definition of moral dilemmas.	140
4.8	Matched moral reasons and inconsistency.	142
4.9	An answer to Williams' challenge.	145
4.10	An answer to Wiggins' challenge: An unsuccessful attempt.	149
4.11	Two answers to Wiggins' strengthened challenge.	151

Chapter Five: Minimalism and the Location Problem	155
5.1 Platonism, platonism and queer facts.	158
5.2 Minimalist platonism and the epistemological challenge.	161
5.3 Minimalism and counterfactual mind-dependence.	165
5.4 Minimalism and normativity – rationalism.	170
5.5 Minimalist platonism and explanatory naturalism.	171
5.6 Explaining the usefulness of a concept truth in morals.	176
5.7 Forms of moral relativism.	184
5.8 Moral relativism and the function of moral truth.	187
5.9 Moral minimalism and the function of truth in morals.	194
Chapter Six: Minimalism, Supervenience and Naturalism	197
6.1 Disputational supervenience and Cognitive Command.	199
6.2 Actual substantive moral-on-natural supervenience.	203
6.3 Explaining substantive moral-on-natural supervenience.	204
6.4 Moral minimalism and substantive moral-on-natural supervenience.	208
6.5 Weak inter-world moral-on-natural supervenience (McFetridge).	207
6.6 Strong inter-world moral-on-natural supervenience (Jackson).	219
6.7 Substantive naturalism and disputational supervenience.	224
6.8 Analytical naturalism, disputational supervenience and open questions.	228
6.9 The prospects for analytical naturalism.	238
6.10 Ontological naturalism and disputational supervenience.	242
Bibliography	247

Acknowledgements

This thesis has been written while being matriculated as a Ph.D. student at the University of St. Andrews in Scotland and while having the benefit of a generous scholarship from the Danish Research Academy. My supervisor was Professor Crispin Wright.

Many thanks are due to fellow students and members of staff for comments on chapters of the thesis and papers held in St. Andrews, at Raasay and in Århus. I wish to thank especially members of the 'Southgait reading group' for many challenging evening discussions over three years in St. Andrews: Lars Gundersen, Patrick Greenough, Jesper Kallestrup, Paul Markwich, Duncan Prichard and Svend Rosenkrantz. Other fellow students I wish to thank especially are Jakob Hohwy, Bjørn Jespersen, and Joanna Tedziagolska.

A special debt is also owed to members of my review-board at the University of St. Andrews. Garrett Cullity, Stewart Shapiro and John Skorupski for reading many early versions of chapters and for patiently giving invaluable advice. And thanks also goes to Paul Boghossian, John Broome, John Divers, Hans Fink and Alex Miller for important comments on chapters of the thesis or papers held on various occasions. A special thanks also goes to Bob Hale for generously and helpfully commenting on chapters of the thesis, but especially – as my M.litt. supervisor in St. Andrews – for encouraging me and preparing me for the Ph.D. years.

The thesis itself will leave no reader in doubt about my *philosophical* debt to Crispin Wright. But, my debt is much deeper than this. His inspirational, patient, generous and supremely competent supervision has played a very large role in shaping the thesis and in inspiring the work effort behind it. If I have learned one thing in particular from him, it is never to shy away from the complexity of a subject, but to seek clarity and rigour without compromise – never to rest on the laurels of insights already gained, but always to challenge the assumptions on which it was based.

Århus
September 2000

Lars Binderup

Moral Minimalism

Introduction

At one time it was usual to say that we do not call ethical statements 'true' or 'false', and from this many consequences for ethics were meant to flow. But the question is not whether these words are in practice applied to ethical statements, but whether, if they were so applied, the point of doing so would be the same as the point of applying them to statements of other kinds, and, if not, in what ways it would be different. (*Michael Dummett*).¹

Many moral philosophers in the 20th century have had the intuition that offering an opinion about whether something is morally right or good is not the same kind of thing as offering an opinion about whether, say, the Earth is round or whether this table is one metre high. It has been common to think of the latter as questions about which there are 'genuine facts of the matter'. That is, facts *existing independently* of the judging subjects and facts that make true the statements that succeed in representing (*in a substantial sense of representing*) them correctly. By contrast, many philosophers have held that the truth or falsity of moral judgements is somehow crucially dependent on the subjects involved – there is in some sense no mirroring of *independently* existing facts going on here. In short, non-objectivism and anti-realism about morals have been a popular stance to occupy with respect to moral discourse – and indeed, normative and evaluative discourse in general.

The following thesis is concerned with formulating a new anti-realist and non-objectivist position – to be named *moral minimalism*.² Thus, it will be concerned with explicating the just described intuition about a discontinuity between ethical discourse and, for example, scientific discourse and everyday discourse about the immediate natural surroundings. The central question will be how to give coherent expression to this discontinuity. And here, as it will become evident, the main challenge is to show how to give an account of the discontinuity which respects the fact that we as a matter of fact aim at truth (or at least correctness) when taking part in moral

¹ Dummett (1978), p. 3.

² Throughout the thesis, I shall use the words 'ethical' and 'moral' interchangeably.

discourse. Consider here, for instance, the fact that we use concepts like those of *truth* and *belief* in moral contexts.

Since the focus is therefore on moral antirealism, there will be relatively little on modern attempts to formulate *realist* positions in the following – whether these focus on discontinuity between morality and science, as for instance the various forms of Kant-inspired moral *rationalism*, or on continuity, as various kinds of moral *naturalism* do. The realist positions will, though, have cameo-parts in the following, since it *is* my ambition to place these positions on the dialectical map together with moral minimalism.

There have previously been two main anti-realist paradigms in analytical moral philosophy. Philosophers like Ayer, Stevenson, Hare, Blackburn and Gibbard – to mention some – have attempted to accommodate the antirealist intuition by giving a non-cognitivist and expressivist account of moral discourse. Expressivists hold that when I utter a moral sentence, I am not asserting anything. Rather I am giving expression to a (non-cognitive) attitude of one sort or another. Thus, the expressivist tries to accommodate the antirealist intuition by arguing that moral judgements *merely masquerade as*, or mimic, judgements in assertoric (and truth-apt) discourse and, therefore, that the relation between moral judgements and the world is not one of representation, but rather of projection. There are no moral facts 'existing independently out there in the world'. In a sense, we – or rather our non-cognitive natures – put them there ourselves.³

Philosophers working within the second main – error-theoretic – paradigm have tried to make room for the antirealist intuition by claiming that, though moral judgements *pace* the expressivist are genuinely truth-apt, they are all false. Ethical statements purport to depict states of affairs of a queer kind – states of affairs that can mysteriously cause me to be motivated to act. But, the world simply does not contain such things. And so all statements of ethics – implicitly claiming, or so argues the error-theorist, that there are such states of affairs in the world – are false. In other words, moral discourse is like discourse about witches and phlogiston in that it is

³ A related position in another tradition in philosophy is, of course, that of existentialism as advocated by philosophers like Sartre and Camus.

based on a false presupposition. Of course, Mackie is the main figure in this anti-realist tradition.⁴

Both non-cognitivists and error-theorists have encountered serious problems for their different accounts. The question that poses itself is, therefore, whether a philosopher who advocates both cognitivism and a success-theory about moral discourse can accommodate the intuition. Is it coherent to hold that moral judgements are apt for truth, and that some are true, and yet hold that they do not aim at representing independent states of affairs? If one wants to defend a view of this type, one is obviously committed to concepts of truth, assertion and judgement (and so on) that do not incorporate representationality in any substantial sense. In other words, one has to defend the claim that the concepts of truth, assertion and judgement (and so on) are not exclusively the property of realists.

My thesis is an attempt to defend the novel *minimalist* position in meta-ethics – *cognitivist*, yet *antirealist* – that does exactly that. Thus, my main contention will be, on the one hand, that moral discourse displays an inter-subjective discipline and a syntax that is sufficient for regarding it as dealing in genuinely truth-apt contents and, on the other hand, that moral discourse is not genuinely representational – that there is no aim at correspondence to a reality existing independently of humans implicit in the making of an moral judgement. Indeed, the claim to be defended is that moral discourse lacks *all* realism or objectivity-relevant features. My thesis is that we can have *truth without realism and (in a certain sense) objectivity* in moral discourse.⁵

Importantly, I hasten to add, this thesis is, for reasons that will become clear already in the first chapter, a thesis about moral discourse in a restricted sense of this word. That is, I am claiming merely that a minimalist account is coherent and defensible when advanced as a theory about *discourse about moral principles*. A moral principle – I shall also call it a moral linking-principle – is a conditional with a non-moral description as the antecedent and a moral description as the consequent. There is, of course, a broader notion of moral discourse – more in line with the everyday notion of a moral discussion – that includes singular moral judgements, i.e. moral judgements about particular situations, persons, actions and so on.

⁴ See Mackie (1977).

The real challenge to the defender of moral minimalism – apart from providing a workable anti-realist notion of truth – is to get a clearer grip on the notions of 'correspondence' and 'objectivity'. How do we give an account of them that makes them more than just suggestive, but out-of-date, realist metaphors? In my defence of moral minimalism, I adopt Crispin Wright's framework for realism-debate as laid out in his *Truth and Objectivity*.⁶ The most relevant mark of objectivity and aptness for genuine representation in this framework when considering moral discourse, I argue, is the Cognitive Command test. Failure to pass the test for some discourse is constituted by there being no a priori guarantee that any dispute formulated within the discourse must involve a cognitive shortcoming. This test is arguably the weakest test of a discourse's status with respect to realism and objectivity. Thus, more precisely, my aim is to argue that moral discourse about principles is apt for (minimal) truth, yet fails this particular minimal test of objectivity.

The thesis is structured, accordingly, as a defence of moral minimalism – the claim that moral discourse is truth-apt, but fails the further Cognitive Command test for objectivity. I introduce the position – and the key concepts of Wright's novel framework for debates about objectivity and realism – in more detail in *the first chapter* and I, here, defend it against the first objection in the literature to the effect that it is an inconsistent position. The *second chapter* deals with a second objection to the minimalist framework for realism-debates – an objection based on an agreed feature of moral beliefs, namely motivational internalism: Having a moral belief, pertaining to oneself, tends to motivate. More precisely, the objection is that the minimalist's notion of moral belief – that follows trivially on his cognitivism – cannot be a *genuine* notion of belief because of this motivational feature. However, the defence of a distinction between merely minimal beliefs – or *commitments* – and robust beliefs will help to answer this objection for the minimalist, it is argued.

I then show – in *the third chapter* – how the minimalist position accounts for, shows itself consistent with, what may be called the phenomenon of radical moral disagreement. I argue that intuitions about the possibility of radical moral disagreement is best characterised exactly as the lack of an a priori guarantee that all

⁵ The qualification in the brackets is necessary because some would use the word 'objective' in such a way that truth-aptness implies objectivity. Of course, this only goes to show that a clarification of notions like objectivity and realism is called for.

moral disputes involve a cognitive shortcoming – i.e. as failure of Cognitive Command. In *the fourth chapter*, I show how minimalism is also consistent with the phenomenon of moral dilemmas. This crucially involves characterising moral dilemmas in a way different from many other writers on the subject. In *chapter five*, I then address the question how moral states of affairs – on the minimalist's account – can fit into our overall ontology. This includes considering how a naturalistic (in a non-substantive, methodological, sense) account of morality is consistent with the minimalist account. And it includes arguing that the minimalist account need not become relativistic – in the traditional sense of the word – although it can perhaps be seen also as a coherent *reformulation* of the traditional relativist position. The final *chapter six*, then, addresses the issue of whether the phenomenon of substantive moral-on-natural supervenience – once properly characterised – poses a threat to the stability of the minimalist position. Likewise, the question is investigated whether the truth of some form of (substantive) naturalism – in both analytical and ontological versions – could pose such a threat. The discussion here will turn on whether either of these phenomena impose what I shall call moral-on-natural *disputational supervenience*.⁷

⁶ See Wright (1992).

⁷ For a more thorough description of the content of the individual chapters, I refer here to the opening section of each chapter.

Chapter One

The Stability of Moral Minimalism

The following is an attempt to defend a theory about moral truth and objectivity – a form of moral anti-realism – that I shall call moral minimalism. The point of this first chapter is to identify the position and to take the first steps towards showing – against objectors in the literature – that it is a stable and coherent position.

The structure of the chapter is as follows. The first five sections outline in a rather fast pace the minimalist position. These sections introduce the basic concepts of the framework for debates about realism and antirealism that I shall be working within for the remainder of the thesis, i.e. the framework suggested by Crispin Wright in his *Truth and Objectivity*. I distinguish my minimalism from others on the market (section 1). I, then, introduce the notion of minimal truth and say some more about realism-debates as reconceived by the minimalist (section 2). I go on to argue that moral discourse is minimally truth-apt (section 3). This includes saying something about the kind of assertoric discipline found in morals. I, further, argue that merely minimal moral truth can be identified as – what Crispin Wright has named – superassertibility given (among other things) that moral truth is epistemically constrained (section 4). Finally, I attempt to show that moral truth, given an important restriction to discourse about moral *principles*, is epistemically constrained (section 5).

Having introduced the basic framework and the key concepts and summarised it (section 6), I then (section 7) introduce in more detail the realism-relevant *Cognitive Command Constraint* – as this is defined in the work of Wright. The claim – distinctive of the moral minimalist – is made that moral discourse *fails* to pass the test. I explore (section 8) the relation between this feature of a discourse and the feature of having an epistemically constrained notion of truth governing a discourse.

It has been argued in the literature that the minimalist position is fundamentally incoherent – that *any* truth-apt discourse trivially will pass the test. I present this objection (sections 9 and 10) as it has been advanced – in different versions – by Timothy Williamson and Stewart Shapiro and Bill Tascheck. Finally, I answer the objections (sections 11, 12 and 13). In the answer, I point out that the

claim that a discourse fails to exhibit Cognitive Command amounts to the claim that one ought to *remain agnostic about the a priori status* of the truth that any dispute formulated within the discourse must involve a cognitive shortcoming. This answer involves arguing that an intuitionistic logic should be adopted in moral discourse, arguing that moral indeterminacy ought to be characterised neither as merely epistemic, nor as involving truth-value gaps and arguing that there is no direct implication from some claim's failure to be a priori knowable to anything about the *modal* status of the claim.

Having thus, successfully I will contend, defended the basic consistency of the moral minimalist position, I declare myself ready to see how it fares in the remaining five chapters with respect to various important meta-ethical problems.

1. Moral minimalism and deflationism.

First, why did I choose the name 'minimalism' for the position I am defending? The term has gained wide currency in realism-anti-realism debates in recent years, but it is used in more than one way. Importantly, I shall not use 'minimalism' in the way it has most often been used by deflationists or redundancy-theorists or disquotationalists about truth.⁸ Rather, I shall use the term as Crispin Wright uses it in his *Truth and Objectivity*. Let me briefly point to two related differences between deflationist minimalism and what I call minimalism (i.e. Wright's minimalism) in order to clarify and motivate my use of the term.

The deflationist's central claim is the negative claim that there is no substantial property of truth – no property whose essence we may hope to discover through philosophical analysis.⁹ But, he argues, our concept of truth is not a concept of such a property anyway. Rather – and this is the deflationist's positive account of truth – it is a handy device for performing certain linguistic tasks. It allows us to endorse batches of sentences simultaneously as in 'Everything Horwich says is true'. It allows us to endorse a sentence without specifying its content as in 'Einstein's theory is true'. And

⁸ The tradition is often thought to stem from Ramsey. Current prominent representatives are, of course, Paul Horwich and Hartry Field.

⁹ Horwich (1990), p. 1-2.

it allows us to disquote – to affirm things at the meta-linguistic level that we could have asserted at the object-language level.

If we did not want to facilitate assertion in these ways, we could – according to the deflationist – do away with a notion of truth altogether. The deflationist sometimes sums up his position with the claim that the Disquotational scheme is a full explanation of the meaning of ‘is true’:

(DS) ‘p’ is true if and only if p

Against this proposal, however, Wright has argued convincingly that the deflationist has failed to show that truth is not a substantial property.¹⁰ In fact, (DS) – the role of which is ironically emphasised by the deflationist – gives us exactly a *prima facie* reason to believe the opposite.

To see this, consider, first of all, that the concept of truth is a normative concept. If I claim that some proposition is true, I am endorsing it. More precisely, I am recommending the acceptance of the proposition, because it meets a certain (epistemic) standard or norm. The question to ask the deflationist is now what this ‘living up to a standard’ *could* be on his account. In particular, he appears to be faced with a dilemma.

On the first horn of the dilemma, the deflationist *cannot* simply claim that meeting this standard is meeting a distinctive standard of truth – that complying with this norm is something irreducibly different from being epistemically justified. If our concept of truth registers a proposition’s compliance with such a distinctive truth-norm, that is, it is no longer plausible to claim that it is not the concept of a substantial property – and, thus, the deflationist position disappears from the map.

So, the deflationist appears to be forced on to the second horn of the dilemma. He appears to be committed to holding that calling something true is to recommend it as complying with a norm of epistemic justification – for instance, that of (epistemically) warranted assertibility. But, then the following line of argument will cause trouble for the deflationist. The argument concludes that, though the concepts of truth and warranted assertibility have the same normative force – one cannot aim at warranted assertibility without at the same time aiming at truth, and vice versa – they

¹⁰ Wright (1992), p. 12-24 and Wright (1998b), p. 209-219.

may potentially *differ in extension*. To see how this conclusion is reached, notice that the following is a consequence of (DS) together with the plausible assumption that all contents have significant negations:¹¹

'It is not the case that *p*' is true \leftrightarrow it is not the case that '*p*' is true

Thus, the truth-predicate allows negation to move in and out of quotation-marks. But, notice also that this does not hold true of the warranted assertibility predicate. The following is *not* true a priori:

'It is not the case that *p*' is warrantably assertible \leftrightarrow it is not the case that '*p*' is warrantably
assertible

In particular, in cases where our state of information is neutral with respect to *p*, it is not the case that the fact that we have no warrant for asserting *p* means that we have a warrant for asserting its negation.

So, truth and warranted assertibility are two predicates that can come apart in extension, because they function differently with respect to negation. But, then it must be concluded that the norm of truth and the norm of warranted assertibility – though sharing normative force – are different norms. And indeed this conclusion cannot come as a big surprise. We intuitively accept that to criticise someone for having no warrant for asserting something is something quite different from criticising someone for having a false belief.

But, then the deflationist seems forced back on to the first horn of the dilemma – that is, claiming that the truth-norm is a distinct norm. Therefore, he also appears to be forced to give up his central claim that truth is not a substantial property. Deflationism inflates – to use Wright's term – and flies off the dialectical map.

Once this has been recognised, it follows that there is a second important difference between deflationist minimalism and Wright's minimalism. The deflationist

¹¹ The reasoning being as follows: One instance of (DS) – substituting not-*p* for *p* – is:

'It is not the case that *p*' is true \leftrightarrow it is not the case that *p*

Contraposing the two conditionals of (DS) gives us:

It is not the case that '*p*' is true \leftrightarrow it is not the case that *p*

But, then given the transitivity of the bi-conditional, we have:

'It is not the case that *p*' is true \leftrightarrow It is not the case that '*p*' is true

thinks that once truth has been deflated, it can no longer serve as the focus of serious metaphysical debate – realism-anti-realism debate. If truth is no substantial property, then there is no point in having a philosophical debate over its constitution. There are no longer any interesting ways of comparing the truth-properties in different areas of discourse – e.g. ethics, mathematics and science – as it is done in the traditional realism-debates. In other words, the deflationist becomes a quietist. By contrast, Wright's minimalist still believes that realism-anti-realism debates are possible. There is a genuine property of truth and, therefore, there are potentially interesting philosophical comparisons to be made between different discourses with respect to their degree of objectivity. Wright-minimalism is minimalistic in a non-quietist way.

What, further, is distinctive of Wright's form of minimalism is that the debate over objectivity and truth no longer is to be construed as a debate over whether truth (and the related notions of assertion, belief etc.) applies within a given domain. There is no longer any suggestion that the appearance of truth-aptness, assertoric status, belief-states and so on is anything other than genuine or that all statements about the subject-matter are false due to a false presupposition – as matters were conceived by traditional anti-realists in the expressivist and error-theoretic traditions. Instead, in the minimalist's framework, truth becomes a metaphysically lightweight property that can apply in almost any (indicative) area of discourse. Thus, realists *and* antirealists can accept that the truth-predicate applies in any domain that they otherwise disagree about. The question of truth-aptness is taken out of the equation – so to speak – and realism debates become, instead, debates over what *other* realism-relevant properties the truth-predicate has in various domains. I shall in these opening sections spell out the way of reshaping the realism-debates that is implicit in this proposal.

Two comments about the use of the word 'minimalism' in the rest of this thesis can now be made. First, 'minimalism' is always to be understood as referring to Wright's non-deflationary form of minimalism.¹² Secondly, when I am referring to moral minimalism, I am *not* just referring to any position that employs Wright's framework for realism-debates – any position that uses a minimal notion of truth as a

¹² It may be argued that 'minimalism' is a word that has already been appropriated by the deflationists – and hence that Wright-minimalism ought find a new name for itself. However, if the above quick argument is sound – the demonstration that (DS) contains the means to show that any deflationary account of truth is unstable – then it seems that the name 'minimalism' is soon going to be available for recycling.

common ground between the realist and the antirealist and then proceeds to point to the presence or absence of further realism-relevant properties. Rather, a moral minimalist is someone who both conceives of realism-debates as Wright does *and* argues that moral truth fails to possess any further realism-relevant properties – that moral truth is *merely minimal*. So, 'moral minimalism' – the position I shall now go on to outline and, later, defend – is in the following short for *mere* minimalism about morals.

2. *Minimal truth and the debates about moral objectivity and realism.*

The central minimalist claim about *truth* is that all it takes to be a concept of truth is to satisfy the *basic platitudes* surrounding the concept. That is, the set of basic a priori and necessary truths involving the concept that any competent user of the concept must – at least on reflection – recognize. So, the minimalist's idea is not – and here there is a similarity with the deflationist's central negative claim – that we can give a reductive definition of truth. The minimalist is not out to specify the necessary and sufficient conditions for being true – in the manner of traditional correspondence, coherence or pragmatist theories of truth. Rather, the aim is to give a platitude-based philosophical account of truth. The strategy is to compile critically a list of all platitudes involving the concept of truth – to construct an analytical theory of truth – and then to claim that any candidate for being a truth-predicate which satisfies this theory is successful.

So, first, what are the platitudes surrounding our concept of truth? Here is a possible list. (As will be evident there are entailments between these platitudes that make some of them, strictly speaking, redundant):¹³

1) Truth is *transparent*. To assert is to present as true. This is, of course, equivalent to the Equivalence Schema: (ES) It is true that p if and only if p. And together with the convention that 'p' says that p, this platitude entails the Disquotational Scheme employed above: (DS) 'p' is true if and only if p.

2) Truth is *timeless*. If a proposition is true at one point in time, then it is always true and always was true.

¹³ Wright has suggested lists in Wright (1998b), p. 21-22, (1992), p. 34, (1995), p. 7-8.

- 3) Truth is *absolute*. Truth does not come in degrees, but is always completely present or completely absent.
- 4) Truth-aptness is *preserved under embedding*. If a proposition is truth-apt, then so is it when it is embedded in a negation, disjunction, conjunction, etc.
- 5) Truth *differs from justification*. A proposition can be justified and false and it can be un-justified and true. As we saw briefly above, this is a consequence of the (DS) (or (ES)) together with the platitude about embedding and the observation that we may sometimes be in a neutral state of information with respect to a proposition.
- 6) Truth may sometimes be *opaque* – beyond our ken, for now or forever.
- 7) To be true is – in a trivial, non-substantial sense – to *correspond* to reality, to say things as they are, to fit the facts etc.

This platitude-based minimalist approach to truth is *metaphysically lightweight* in the sense that it does not attempt to capture any deep realist intuitions – there is no attempt to portray truth as being some substantial correspondence-relation between thought or language and the world. Thus, the idea is that even the anti-realist can accept that a discourse, whose realism-status he otherwise disputes, is apt for minimal truth as defined in this way by platitudes – exactly because they are *platitudes*. In other words, the hope is that this lightweight concept of truth can form a common ground for the realist and the antirealist.

Notice now that the minimalist about truth, on the one hand, can claim that – since all concepts of truth must satisfy these platitudes – truth is the *same* across different domains. But notice also, on the other hand, that minimalism makes possible a form of *pluralism about truth*. Not in the sense that the word 'truth' becomes ambiguous, but rather in the sense that we can have different – though formally uniform – concepts of truth operating in different areas of discourse. That is, different predicates can satisfy the analytical theory when we consider different domains of discourse – and the property of truth can as a consequence be differently *constituted* or *realised* in different domains. Thus, for example, truth in ethics and truth in physics – though formally uniform – may be differently constituted.

This, in turn, opens up the possibility of reshaping the traditional realism-debates rather than abandoning them as the quietist deflationist suggests. It becomes possible to compare the claims to objectivity or realism made on behalf of different discourses by focusing on the different truth-concepts and truth-realiser in the

discourses and by asking how they differ with respect to *further* realism- or objectivity-relevant properties. Thus reshaped, realism-debates will revolve around the question whether the discourse can sustain a more *robust* truth-predicate of some kind – a concept that captures different further realist intuitions. And, the anti-realist's *merely minimal* truth-predicate – the kind of truth-concept that captures no further realist intuitions – can then function as the default position or common ground in the debates.

Wright suggests in *Truth and Objectivity* that there are at least four interesting realism-relevant features that truth-realiser within different discourses may possess – four cruxes for realism-antirealism debates. Let me introduce them briefly. The first two cruxes concern the relation between the concepts of truth and superassertibility – the latter being the anti-realist notion of truth as *de facto* durable justification (I shall discuss it later in this chapter). A first question – what may be called the Dummettian debate – is whether these concepts within a given discourse are *co-extensive*. If they are, then truth is *epistemically constrained* – that is, it is a priori that all truths are knowable. If they are not, then truth is *potentially evidence-transcendent*. The latter feature is, of course, the feature that captures a strongly realist intuition: that truth may be 'out there' absolutely independently of us, outside our ken as a matter of principle.

But, even if truth is epistemically constrained with respect to a given domain – even if truth is therefore co-extensive with best opinion – there is still a further way in which truth and superassertibility may come apart, giving rise to what Wright calls a *Euthyphronic* debate (about truth). That is, if a judgement's superassertibility *constitutes* its truth – if judgements about the domain are true *because* they are what best opinion says and not vice versa – then truth in the discourse is Euthyphronic. If, on the other hand, best opinion *merely infallibly tracks* truth – if the judgements' being superassertible is explained by their being true – then truth in the discourse is *Socratic*. Again, the latter feature captures the realist intuition that it is the world, and not us, that make judgements true.

According to Wright, there are two further cruxes in the debates about realism. They relate to the notions of *correspondence* and *fact* as they typically feature in realistic thinking. There is, first, the question whether or not practitioners of a discourse function in a *genuinely representational mode* when they form opinions.

This feature links up with traditional realist intuitions about *convergence* in the opinions of ideally rational deliberators. When conceived of adequately, the question here, according to Wright, becomes whether it is a priori that all disagreements within a given discourse must be down to a *cognitive* shortcoming on the part of one or both disputants. If a discourse has this objectivity-relevant feature, then it exhibits what Wright calls *Cognitive Command*. I shall say much more about this feature later in this chapter (and indeed in the remainder of this thesis).

Finally, there is a realism-debate to be had with respect to the kind of fact or state of affairs that statements formulated in a given discourse correspond to if true. Now, it is a mere platitude that, in so far as a discourse is apt for minimal truth, there are facts corresponding to true statements within the discourse – viz. the seventh platitude above. So, even the strong anti-realist about some domain – the defender of merely minimal truth in a discourse – allows for our speaking of facts or states of affairs as long as the discourse in question is minimally truth-apt. But, a further realist intuition has it that 'real' facts are strongly independent of us – sitting 'out there' in the world and playing an independent role in the order of things. The question is, therefore, how to give adequate expression to this intuition. It is not going to be sufficient simply to say that realistic facts can play genuinely explanatory roles – can feature in explanations of some kind – whereas minimal facts cannot, since even minimal facts can feature in perfectly good explanations. For instance, the fact that something is funny – discourse about comedy is, of course, a prime candidate for a merely minimally truth-apt discourse – can explain my laughing, my judgement that something is funny and perhaps my buying the video.

So, the realist intuition about facts will have to be cashed out in a different way. Wright suggests that what we should look for are facts with a *Wide Cosmological Role*.¹⁴ The realist's facts differ from merely minimal facts, not in their playing an explanatory role, but in virtue of the wide range of different types of explanations in which they can figure. The minimal fact of something's being funny above figured in different explanations, but, crucially, they were all explanations of states of affairs that either were propositional attitudes of someone with the adequate mastery of the concept of funny – like the fact that I judged something funny – or

¹⁴ See e.g. Wright (1992), p. 196.

were states of affairs – like the fact that a video-tape was bought – that had come about as a causal result of someone's having that propositional attitude. Thus, merely minimal facts only have explanatory potential when they work via the mental states of subjects – they are *explanatorily lazy*, so to speak.

Wright's idea is, therefore, that we may regard a given discourse as apt for a more robust notion of truth, if the facts corresponding to true statements of the discourse are apt to figure in other types of explanations – explanations of facts that are not, and do not go through, propositional attitudes that take those very facts as their object. For instance, the fact that it is raining may figure in explanations of the wetness of the grass, in my feeling cold or in my slipping on the wet grass – i.e. explanations that do not involve propositional attitudes taking the fact that it is raining as object. Hence meteorological facts have Wide Cosmological Role. This feature of having wider explanatory potential, thus, captures the sense in which 'real' facts are *independent* of human subjects.

In short, a minimalist approach to truth provides an opportunity to restructure and tidy up the debates about realism. The applicability of a notion of truth within a given discourse is no longer an issue between the realist and anti-realist. The question is rather: What kind of truth? And, the minimalist's framework for realism-debates allows us to see that there are many different such debates – corresponding to various different realist intuitions – that must be kept apart and that different cruxes may be more or less interesting and controversial with respect to different discourses.

3. *Minimal truth-aptitude, discipline and moral discourse.*

In order to provide such a metaphysically light-weight framework for realism-debates, the minimalist has to tailor his conception of *truth-aptness* to fit. His simple idea is now that the sentences of a discourse will be truth-apt, if they can sustain a minimal truth-predicate. But, the minimalist goes on, any *assertoric* discourse – defined in a plausible way – will be able to sustain such a predicate. We should, that is, recognise that there is an a priori and necessary connection between truth-aptness and assertoric status of discourses – to assert is to present as true and vice versa. But, then the question immediately becomes: How should the minimalist define 'assertoric'? Clearly, he will want sentences with assertoric content to have all and only the features

necessary for sustaining a minimalist truth-predicate. This in turn is, as argued by Wright, fulfilled by setting up *two conditions for being assertoric*.

First of all, sentences with assertoric content must display the right kind of surface *syntax*. A sentence from an assertoric discourse must crucially be able to feature as the antecedent in a conditional in order to allow the Disquotational Schema to be partially definitional of a truth-predicate for the discourse. Likewise, assertoric discourse must allow that its sentences can be negated – as well as be connected by other logical operators. In short, a discourse must – in order to be assertoric – have all the syntactic features which enable its sentences to sustain standard truth-preserving inferential practices. Finally, the syntax of an assertoric discourse must allow that its sentences can be embedded in propositional attitudes – in particular, belief contexts. These criteria, we may notice, are quite easy to fulfil – any indicative discourse will satisfy them.

The second requirement concerns *discipline*. Any assertoric discourse has to be governed by publicly acknowledged standards of warranted assertibility. The point of this requirement is to secure that the unit sentences of the discourse in question have a determinate sense at all. If there is no general consensus about what conditions justify the use of the sentences – if there are no established patterns of use of the sentences – there is simply no reason to hold that the sentences have fixed contents. An example of a discourse which does *not* live up to this requirement may be a private language in Wittgenstein's sense. Echoing his famous arguments, if it is not possible to distinguish between *seeming* right and *being* right within a discourse, then there will be no sense in which the sentences can be said to have determinate contents.

It is important to notice that fulfilment of these two constraints on assertoric discourses is to be read off from the surface character of a discourse. The minimalist will *not* draw a distinction between a discourse's *appearing* to be assertoric and a discourse's *being genuinely* assertoric. All that it takes for a discourse to be assertoric is that the sentences of the discourse on the surface behave in accordance with the two constraints, since that is what it takes for a discourse to be able to sustain a minimalist truth-predicate.¹⁵

¹⁵ This feature of the minimalist's account of morals is what gives it a major advantage over the expressivist's account. The expressivists have to explain why moral discourse behaves like an assertoric discourse while not being genuinely assertoric. In particular, he is saddled with the

Now, if we accept this minimalist approach to truth and truth-aptness, the first question to ask when engaging in debate about truth and objectivity in morals is therefore: Are moral sentences truth-apt? And to answer this, we must ask whether moral discourse displays the appropriate syntax and discipline. The former is easy to establish. Moral discourse certainly does appear to have the appropriate surface syntactic features. That is, there seems to be no objection to constructions like the following: ‘I believe that/wonder whether it is wrong to tell lies’, ‘If it is wrong to tell lies, it is wrong to make others tell lies’ and ‘It is not the case that it is wrong to tell lies’.

So, the important question becomes whether moral discourse exhibits assertoric discipline. What is needed is an argument for the claim that there are acknowledged standards of warranted assertibility governing moral discourse – something that imposes a distinction between a statement’s *seeming* true to me and its *being* true. Now, since we are aiming at a characterisation of moral discourse as apt for merely minimal truth, we cannot assume that some domain of moral properties or objects existing independently ‘out there’ imposes this discipline – this distinction between seeming right and being right. It would seem that we must defend the view that there is a discipline governing moral discourses which is *internal* to the community of speakers – and hence the epistemology of the discourse must be dialogical or intersubjective rather than one of correspondence.¹⁶ But, what kind of discipline could that be? To answer this question, I must make a short excursion into moral epistemology. The leading question must be: How is it possible go wrong in making a moral verdict?

First of all, it must be noticed that moral discourse is closely tied up with affective and motivational responses of human beings – for instance, feelings of anger, resentment, disdain, self-loathing and guilt and motivations to act in the form of impulses, repulsions and aversions. To use John Skorupski’s term, moral judgements are ‘affectively grounded’.¹⁷ This is not to say that all genuine moral judgements are accompanied by strong feelings and emotions. But it is the dual claim that in a sense

problem of giving an account of inferences involving mixed statements – statements only some of which are genuinely assertoric according to the expressivist (the so-called the Frege-Geach-Hale problem).

¹⁶ I borrow these terms from Skorupski (1997). Importantly, though, moral propositions are, according to Skorupski and *pace* the minimalist, not merely hermeneutically disciplined.

the whole practice of moralising rests on our spontaneous affective moral reactions and that moral judgements are about what it is reasonable to feel and what motivations (actions) are justifiable. Moral judgements are in this way *grounded* in our affective natures and they *guide* our emotional and practical lives.

Importantly, this is not to say that our moral feelings and motivations are somehow given, in advance and once and for all, before we start to reflect on and discuss morals. They are not somehow a *fixed* natural basis on which moral discourse is erected. Rather, our moral emotions and motivations – though obviously made possible by our natural constitution – are shaped by our moral upbringing. What we are affectively prompted to judge right or good is not pre-programmed by nature, but passed on to us by parents, teachers and other role-models and perhaps (hopefully) later revised by us as we develop the ability to reflect on our moral norms.

Let me now suggest what is in effect a moderately externalist theory of moral justification. The suggestion is that a speaker's having spontaneous moral feelings and motivations *entitle* – though defeasibly and *prima facie* – the speaker to hold corresponding moral beliefs. The notion of an entitlement employed here is the notion of an epistemic right that one *does not have to earn* (but, of course, *may earn*) through evidential labour, but is there as a default.¹⁷ In other words, it is the idea that the right to have some moral conviction (whether in the shape of a moral principle or as a judgement about a particular situation) exists whenever one has *no reason not* to have this conviction – i.e. whenever no consideration defeats one's default entitlement.

But, why should we believe that moral affective responses – for example, an emotional inclination to reject or praise an particular action – give us epistemic *rights* in the shape of such entitlements. Let me here just make the proposal – inspired by the views of Dretske¹⁹ – that what entitles us to believe that *p* here is that we (or rather, those of us that are 'epistemically responsible') *cannot help believing* that *p*. As Dretske puts it, 'we have a right to accept what we are powerless to reject'. And moral reactions to situations or types of situations, it seems to me, have exactly this feature of irresistibility. If some action prompts a feeling of moral outrage in me, I

¹⁷ Skorupski (1997), p. 353.

¹⁸ I propose to use the term 'entitlement' in the way it is used by for example Dretske. See Dretske (2000), p. 591-93.

cannot prevent this from causing my forming the belief that the action is morally wrong.

However, the idea of such entitlements would be very naïve, if such entitlements were *indefeasible*. If that were the case, then moral discourse would not be disciplined. If, that is, merely feeling something spontaneously were on its own sufficient to make one's moral judgement indefeasibly warranted, there would be no distinction to draw between my merely seeming to have a warrant and my really having one. There would so far precisely be a *lack* of discipline.

Therefore, if this suggestion about moral epistemology is on the right track, then we must look for ways in which the entitlement to an evaluative judgement on the basis of spontaneous and irresistible feelings can be defeated in order to argue that moral discourse is disciplined. Let me, therefore, briefly point to three (interrelated) ways in which such an entitlement to holding true a singular moral judgement – e.g. that *that* act was wrong – might be defeated.

(1) One obvious way in which a person's entitlement to an affectively grounded judgement may be defeated is if the person has *false information* about the non-moral features of the situation that he is judging. Or, if some *reasoning* involved in forming the judgement was somehow flawed – prejudiced, confused, inconsistent.

(2) There is also a very basic way in which a person's affectively grounded entitlement may be defeated. It may fail even to be *intelligible* as a moral judgement. Arguably, there are certain basic conceptual truths that any competent moral judge must comply with in order to be deemed intelligible as a moral judge. There are, for instance, conceptual truths about the *content* of moral judgement – for instance that moral judgement must take account of all persons involved – such that ignoring these while making moral judgements will makes these judgements unintelligible as *moral* judgements. (I shall say more about other moral platitudes in chapter 3). Thus, contravention of moral platitudes can lead to the defeat of singular moral verdicts.

(3) One conceptual truth about morals that arguably plays a very central role is the supervenience platitude (so I will mention it separately from the other platitudes). It is a conceptual truth that if I were to judge that two situations – though exactly alike in all relevant non-moral respects – merit different moral evaluation, I am

¹⁹ Dretske (2000), p. 598-601.

guilty of a form of moral inconsistency. Thus, a third way in which my entitlement to holding true a singular moral judgement may be defeated is if I make contradictory moral judgements about relevantly similar situations. (I shall say more about supervenience – and the related concept of universalisability – in chapter 6).

Thus, it is clear that it is only *some* entitlements based on spontaneous feelings and motivations – those that are made by rational, well-informed judges, are intelligible as *moral* judgements and survive critical examination in the form of a supervenience-test – that survive. Hence, we have seen in rough outline how affectively grounded moral entitlements may be defeated. It is, therefore, now a little clearer why the moral minimalist – pace the expressivists – can claim that affectively grounded discourse can display the kind of discipline required for having assertoric status. An outline of a case for the claim that moral judgements are genuine judgements – displaying as they do both appropriate syntax and discipline – has been drawn.

Let me, however, make a further comment about the moderately externalist theory of moral epistemology just presented. The fact that one may be entitled to hold a moral proposition true without having earned this right through providing positive justification does, of course, not exclude that one may attempt to provide such positive justifications anyway. Arguably, normative moral *philosophy* is exactly such an attempt to provide systematic justifications of our moral judgements.

In this philosophical enterprise, I argue, it is a general requirement that one's singular moral judgements are *justified by inference from universal moral principles*.²⁰ This requirement sets up the familiar dialectic between moral principles and singular moral verdicts which in turn creates further possibilities for defeat. Not only can universal principles falsify singular verdicts. Singular moral judgements may gang up to defeat a moral principle that we have previously internalised or consciously accepted. And, more general principles can fail to cohere with more specific principles. In other words, in normative moral philosophy we seek a *reflective equilibrium* between our singular and our universal moral judgements – to use Rawls' famous expression – and in this process, we may have to revise judgements of both

²⁰ Of course, this conception of what normative moral philosophy is (and ought to be) is not uncontroversial. Thus, I am here declaring myself in disagreement with a moral particularist like Dancy (1993).

kinds. Thus, with this attempt to provide systematic justifications, we add a further level of discipline to our discourse about singular moral judgements in that we demand that our singular moral judgements must cohere with our preferred set of moral principles – our best normative theory.

Finally, it should be noticed that entitlements to judgements about moral *principles* cannot be defeated in *one* of the ways in which an entitlement to a judgement about particular situations can be defeated. That is, they cannot be defeated in the first way mentioned in the list above – i.e. defeat due to false information or ignorance about the particular circumstances about which one is judging morally. So, for the type of moral discourse that shall interest us in the rest of the thesis – discourse about moral principles – possible defeat (and therefore discipline) is a question of lack of intelligibility or of lack of coherence with a body of other moral principles.

4. Superassertibility as truth in morals

The minimalist framework for realism-debates just outlined is the framework that will be adopted in the following thesis, where the aim is to defend the claim that moral discourse – while truth-apt – does not pass *any* further realism-relevant test. In other words, the aim is to defend a radically antirealist – or merely minimalist – position about ethical discourse.

Given this overall aim, the next important question to ask is what kind of truth *could* constitute truth in morals given such a thoroughly antirealist – *merely* minimalist – conception of morals. The minimalist has to come up with a satisfier of the truth-platitudes that is metaphysically neutral in the sense that it has no realist content. One way to do this, I now suggest, is to construct a notion of truth out of the standards of warranted assertibility that govern any assertoric discourse. Intuitively, if our notion of moral truth is in this way a mere projection from the actual assertoric standards of our moral practice, there will be no appeal to any realistic intuitions about correspondence to an independently existing reality in our account of moral

truth. Wright has argued that such a language-game-internal notion of truth – what he names *superassertibility* – can play the role of moral truth.²¹

The intuitive idea behind the notion of superassertibility is that being true – at least in certain areas of discourse – is simply being warrantably assertible (or believable) where the warrant is accessible and, *as a matter of fact*, indefeasible – i.e. where the warrant would survive the best, actually possible, scrutiny.²² A more precise definition is:

A proposition, *p*, is superassertible if and only if there is some accessible state of information in which *p* is warranted and, no matter how *as a matter of fact* this state of information is or can be enlarged upon, *p* will remain warranted.

To investigate the moral minimalist's claim – that moral truth is superassertibility – we must ask whether the notion of superassertibility satisfies the truth-platitudes in the case of moral discourse. I shall argue – closely, but not entirely, following Wright²³ – that it *can* satisfy the platitudes given three plausible further assumptions about moral discourse:

- 1) That moral truth is epistemically constrained so that it is a priori that moral truths are knowable. (For all moral *p*, *p* if and only if *p* is knowable).
- 2) That evidence for moral claims is of timelessly accessible kind – at least with respect to universal moral principles.
- 3) That the right account of moral knowledge and justification²⁴ cannot be strongly externalist – as in extreme or ‘Mad Dog’ reliabilism.²⁵ This excludes the possibility that one can know that *p* while thereby laying oneself open to a charge of irrationality.

I shall argue for the first assumption in the next section. The second assumption is plausible, I argue, since the way we come to justify universal moral judgements is

²¹ See Wright (1993), pp. 411-18, and (1992) pp. 44-61.

²² Notice that there is *no* demand that the evidence for the relevant *p* has to be *conclusive*. It suffices for the justification for *p* to be *inconclusive and in this sense defeasible*. (Conclusive evidence is defeasible evidence too, but with the added feature that if *p* turns out to be false, then some part of the original justification of *p* will also turn out to be false – e.g. if my justified belief that the result of some complex arithmetical calculation is true turns out to be false, then at least one of the assumptions I made when doing the calculations – that I remembered to carry, for example – will also turn out to be false).

²³ See Wright (1998b), Appendix, Illustration II.

²⁴ By including this rejection of strongly externalist theories of moral *justification*, I appear to be deviating from Wright (1998b).

²⁵ For a rejection of this form of reliabilism, see Dretske (2000).

through a priori conceptual reflection – we attempt to square intuitions about particular cases (often imagined) with our set of universal principles in order to find a reflective equilibrium. And the ability to engage successfully in this process of reflection is not dependent on its taking place at a particular point in time. (Of course, this assumption can only be made about discourse about moral *principles*, since evidence about particular acts or situations may be lost as time passes. Thus, there is an important restriction to discourse about moral principles when it is claimed that moral truth is superassertibility).

The third assumption is a plausible general constraint on any theory of knowledge and justification (or entitlement). For example, it cannot be the case that someone knows that p (or is entitled to believe that p) due to some reliable cognitive mechanism, while opening himself to a charge of irrationality by also possessing all things considered evidence against p . (Notice that this constraint on theories of knowledge and justification does not exclude moderate forms of externalism like the one defended for moral discourse above, since this theory would regard an entitlement to believe that p – and hence a claim to know that p – as defeated in this scenario).

Does the notion of moral superassertibility satisfy the seven platitudes about truth we listed in section 2 given these assumptions? We can simplify matters by remembering that platitudes one (about asserting being to present as true), five (about the distinctness of truth and justifiability) and seven (about platitudinous correspondence) are all implicit in the Equivalence Schema:

(E) It is true that p if and only if p

Therefore one important question becomes whether the following equivalence holds true a priori:

(E_s) It is superassertible that p if and only if p

To take the left to right direction first, assume for reductio that p is superassertible, yet false. Given the assumption about epistemic constraint on truth, we can come to know that p is false. But on the other hand, given the superassertibility of p , we can have *a de facto* indefeasible all-things-considered warrant for p . So, given that moral

epistemology is not such that having one bit of evidence excludes having another – the evidence for moral claims is timelessly accessible and, hence, additive – our assumptions lead us to hold that someone can be in a state of information in which he knows that p is false while also having an (all-things-considered) warrant for p . But, it is absurd, given the third assumption, to hold that the claimant can know that p is false – and hence believe that p is false – yet also have (de facto indefeasible) all-things-considered evidence *for* p , since this would open the claimant to a charge of irrationality. So, it is never the case that we have a counterexample to the left to right conditional in (E_s).

The other direction of (E_s) can be proved by reducing to absurdity the assumption that p is true, yet not superassertible – given the assumptions about moral discourse above. If p is true, then p is knowable (given epistemic constraint on moral truth). And further, if someone knows that p , he believes that p . On the other hand, if p is not superassertible, then any warrant for p can be defeated. And, given the account of moral epistemology given above, this will mean that any positive justification as well as any entitlement to believe that p can be defeated. But, arguably it is not rationally possible to sustain the belief that p while its being the case that any entitlement or justification to believe that p can be defeated. Hence, by the third assumption – excluding any sort of epistemology that lays a claimant to knowledge open to a charge of irrationality of this sort – it is not possible to know that p . So, given epistemic constraint, p cannot be true which, absurdly, contradicts our initial assumption. So, neither can there be a counterexample to the right to left conditional of (E_s).

What about the last four platitudes? Well, clearly moral superassertibility is a timeless affair. Given the assumption that evidence for moral claims – about moral *principles* – is timelessly accessible, there is never the possibility that there is some *de facto* indefeasible warrant for p that is available at one time, yet not at another.

Also, moral superassertibility must be absolute. Though warrants have differing strengths or degrees, the question of whether or not some accessible state of information is such that it warrants p and, no matter how it actually is or can be enlarged upon, it continues to warrant p , is an absolute matter – either this type of warrant exists or it does not exist.

Likewise, aptitude for superassertibility is preserved under embedding within logical operators. After all, to be superassertible is just to be durably warrantable. And, since aptitude for warrant *is* preserved under logical embeddings, so must aptness for superassertibility be.

Finally, the superassertibility of some *p* can, of course, as a matter of fact be outside the cognitive reach of a particular subject (or group of subjects). Someone may not have gathered or processed the evidence for *p* yet, and so not at present have a warrant for *p* although a state of information that warrants *p* is actually accessible. However, the fact that some moral *p* is superassertible cannot be beyond the ken of subjects *as a matter of principle*, since *p*'s superassertibility guarantees that there is an *accessible* state of information that warrants *p*. So, moral superassertibility can satisfy a weak version of the platitude about opacity – the claim that the superassertibility of some moral *p* can escape any subject at any time, but not as a matter of principle. But, this is what we would expect, given the assumption that moral truth is epistemically constrained.

So, it is arguable that the concept of superassertibility is a satisfier of the truth-plateitudes in the moral case. However, we have also seen that this is true only under the assumption that moral truth is *epistemically constrained*. I shall now turn to arguing that this is a plausible assumption.

5. *An argument for epistemic constraint on moral truth.*

Might there be in principle unknowable moral truths? Intuitively, at least according to my intuition, this question must be answered negatively. It seems to me that the whole point of moral practice is for us to find guidance on how to act. And the idea of moral truths that cannot even in principle be known by agents, yet are essentially there to guide agents – of truths that cannot as a matter of principle play the role that they are designed to play – is, it seems to me, mysterious.²⁶ But, these intuitions are not likely to convince those who do not share them. So, let me try to argue this point – that moral discourse fails to have the realism-relevant property of being apt for potentially evidence-transcendent truth.

²⁶ Griffin (1996), p. 106 appears to have similar intuitions.

First of all, a remark about the dialectic of the debate over this issue. It is a general feature of the minimalist framework for the realism-debates that it always leaves the onus of proof with the realist. As already mentioned, the antirealist – merely minimalist – position is the default for any discourse pending a positive argument showing that the truth-realist in the domain has some further realism-relevant property. In the debate about epistemic constraint, it is the realist who is making the stronger claim – that there are or could be these unknowable truths *in addition* to the knowable ones. Therefore, it is up to him to show how we can come to know a priori that there are or could be such potentially evidence-transcendent truths. He must show this in some indirect way, since there is, of course, no direct way as the unknowable truths themselves are just that – unknowable.

To see what strategy the realist might adopt to support this strongly realist claim about moral truth, let me first of all look at another discourse about which the claim that there are unknowable truths *does* seem to make sense. Thus, we do seem to be able to tell a story that helps us make sense of the idea that there are truths about what happened at some time in the past even though all evidence about the event have long ceased to exist. This might be made intelligible by reference to three very plausible claims. 1) We can know what happens *now* in our immediate surroundings and what happened in the near past (through perception, testimony and memory). 2) It is an indispensable feature of our language that truth-values are linked as illustrated in this example: "My fridge is empty on 1 July 2000" is true if and only if "My fridge was empty on 1 July 2000" is true in the year 2001.²⁷ 3) It is a platitude that truth is timeless – if a statement is true at some time, it is always true. Now, it may very well be knowably true today (1 July 2000) that my fridge is empty on 1 July 2000, yet also be the case that in the year 2001 there will be no evidence left of this trivial fact. Say that I will have forgotten and didn't record the fact (and, I assume, time-travel is an inconsistent notion). But then, given 2) and 3), "My fridge was empty on 1 July 2000" will be unknowable, but true, in the year 2001.

Thus, we can tell a story lending plausibility to the claim that there are unknowable truths about the past. Now the question is whether something analogous can be done for moral truth. Of course, this can be done trivially for the truth of

²⁷ See Wright (1993), p. 179.

certain singular moral propositions. For instance, propositions about the moral properties of particular acts in the remote past may be unknowably true or false due to permanent lack of evidence about the non-moral character and circumstances of the act. But, arguably there cannot be unknowable *universal* moral truths – true moral *principles* – forever beyond our ken. As already noted (section 3), evidence for moral principles is arguably timelessly available to moral deliberators and, hence, there is no chance of a loss of evidence as was the case with knowledge about the past. So, the realist must look for some other way of explaining why moral truth (about *principles*) could extend beyond our ken. Let me suggest that there are two major schools of thought about morals that might be thought – implicitly – to advance such an explanation: Theism (or divine law moral theory) and utilitarianism.

First, the theist. A version of the theist's story about morality that lends plausibility to the claim about unknowable moral truths may simply be this: God has designed a moral order for the universe, but put human beings in a situation where they cannot, even in principle, come to know about it or parts of it. We can, for example, imagine a possible world in which there is a god and where this god did not hand Moses the ten commandments and did not in any other way let his creatures know about his thoughts on how they ought to act. Quite apart from the immediate strangeness and potential inconsistency of this very idea – at least on the standard conception of God: is it consistent to hold that a *benevolent and omnipotent* being would do something like that? – is this a way of making sense of unknowable moral truths? I shall argue that it is not after introducing the utilitarian story.

Secondly, an act-utilitarian claims that we can identify a single essential objective good in the world – for instance, pleasure, utility.²⁸ He, then, claims that the criterion for an action's²⁹ being (all-things-considered) right is that the action has the

²⁸ I am in the following taking the act-utilitarian as my example, but a similar point can be made about rule-utilitarians. Rule-utilitarians adopt the view that the criterion for rightness for *actions* is that they comply with certain rules – or, in other versions, are motivated in certain ways or spring from certain virtuous dispositions – but that when choosing the *rules* (or motives or virtues), we have to maximise the expected utility of their being adopted (or apply some other function) by the agent or by all agents. The arguments that follow against the act-utilitarian's way of making sense of unknowable moral truths apply also to the rule-utilitarian.

²⁹ Notice that the act-utilitarian does not have to claim that his theory is one about moral *decision-procedures*, about moral *deliberation*, in addition to being about criteria for moral rightness – i.e. being about how to *justify* moral beliefs about rightness. A sophisticated version of act-utilitarianism grants that we may not want to start moralising whenever faced with a set of options for acting – when what is called for is swift or spontaneous action, for instance. Rather, the options available to

best expected consequences for all persons involved. Depending on what function from good to right is chosen, the right is seen as what maximises the good or what maximises the good-given-that-each-person-should-be-secured-a-certain-minimum (a satisficing function) or what maximises the good of the worst off – or something fourth. Now, there appears to be two ways in which such an act-utilitarian theory might yield unknowable truths. It may, first, be that gathering and processing evidence about the long-term expected *non-moral* consequences of some act is in principle impossible. Perhaps calculating the expected consequences on the whole of society far into the future is just so complicated as to be beyond humans. But, there are two *prima facie* problems with telling this story about unknowable moral truths. First, it is not clear that a wise act-utilitarian will want to take anything other than *expected, relatively short-term, consequences for those immediately involved* into consideration. However, this pragmatic observation is hardly going to answer the philosophical question of whether certain inaccessible truths will elude this wise act-utilitarian. But there is a second, more decisive, problem in that this story poses a problem only for attempts to know *singular* moral truths – facts the evidence for which is partly non-moral. And, as argued above, we are here looking for ways in which *universal* moral principles may be in principle unknowable.

So, in order for the act-utilitarian to make sense of unknowable truths in the relevant sense, it has to be *the very function taking us from good to right that is possibly beyond the reach of human subjects*. The idea would have to be that the basic principle – stating that some particular function from good to right is the one to use when justifying beliefs about what is the right thing to do – would be so complex that it was in principle impossible for humans to grasp it. If the act-utilitarian account

the moral deliberator when engaging in utilitarian reflections include options that are *not directly behavioural*. Thus, one option (to calculate the consequences of) might be the attempt to acquire a certain habit of acting in accordance with a certain moral principle or motive in the heat of action. This means that – as a theory about moral decision-procedures – the act-utilitarian might go for an *indirect* (or restrictive or strategic) theory according to which the license to deliberate according to the act-utilitarian principles in the heat of decision-making is restricted. This is to be distinguished from the rule-utilitarian who advances compliance with rules as a *criterion* of rightness. For more about these distinctions, see Pettit (1991b). (A potential problem for the restrictive utilitarian is that there still appears to be a problem about deciding how to act in the heat of the moment, since there is, arguably, always a decision to make about whether this situation is one that calls for spontaneous or habitual reaction or whether utilitarian calculation is called for. Thus, it may be *in principle* impossible for the utilitarian to leave room for spontaneity in action – at least if he is not able to explain why there is no *decision* to be made here).

makes this claim³⁰ – that there is a universal truth about what is right, but that it is unknowable in principle – he *does* appear to have a relevant story making sense of unknowable moral truths.

So, *prima facie*, we have two proposals for making sense of unknowable moral truths – a theist and a utilitarian. I now want to argue that they both fail – and for essentially the same reason. To see this, let us imagine a society existing in a world of which the theist's or the utilitarian's story is true. In other words, imagine that answers to the Moral questions posed by the inhabitants of this possible world are always beyond what they can recognise.³¹ I capitalise 'Moral' here to indicate that the truth-conditions of moral statements like "x ought, all things considered, to ϕ in conditions C" – referring as they do to an unknowable will of God or complex unknowable functions – are potentially unknowable in this world. The question now becomes: How can these people, who cannot even in principle come to know what is the Right thing to do, be able *even to try to do* the Right thing? It would seem that they cannot. We can make no sense of someone's *aiming* at a target, if he has no means whatsoever of determining whether he has been *successful* or not.³² Discourse about Morals would then be *unable to play the role of guiding* the behaviour of its practitioners and hence be pointless. There would be no way to make sense of disputants discussing their moral views and criticising each other morally, since no-one could ever be proved right or wrong.

But now, it also seems clear to me that people in this predicament could still ask intelligible prudential and moral questions: "How do I live a good life?" and "How do I pay due respect to others in my life?". They could for instance do this by, first, attempting to identify a list of goods – understood as things that meet their rational interests. And secondly, by trying to devise – through normative debate – manageable prudential and moral norms to guide them in their attempt to, e.g., maximise the

³⁰ I am not sure that any actual utilitarians make this claim.

³¹ It might be objected here that the moral realist of the theistic kind does not have to claim that God has withheld *all* means of knowing moral truths to make sense of the idea of unknowable universal moral truths. He can claim merely that God has not given us capacities to know a class of very complex moral truths, but that he *has* given us a fair chance to come to know the simpler ones. However, in order for this scenario to help out, the moralisers in this world would have to be *able to tell* when, and when not, they were dealing with a moral fact that they could know about. And, it seems that this ability to distinguish between the knowable and the unknowable moral facts is only possible if the moraliser – absurdly – knows what the unknowable truths are.

goods they enjoy and to ensure that they respect the interests of others – where this is all relative to that list of goods. So, we can imagine that a practice of *moralising* would develop in this society. A practice performing the functions in society that *actual* moral discourse, arguably, performs – i.e. that of co-ordinating and adjusting and improving the set of norms that structure social life and facilitate peaceful co-operation. This form of *morality* (without capitalised first-letter) might in no way resemble the set of norms that God made, but didn't reveal, or the complex functions of the extreme act-utilitarian.

My point here, thus, is that there seems to be absolutely no reason for saying that it is *Morality* rather than *morality* that is what we in the real world call 'morality'. And, of course, it is the nature of the latter that we are trying to determine when we argue about the status of moral discourse in the realism-debates. Crucially, *morality* need involve no commitment to potentially evidence-transcendent truths.

There may, of course, be ways for the semantic moral realist – the believer in unknowable moral truths – to carry the onus of proof other than via defending versions of theism and utilitarianism. But, to conclude for now, *so far* we have seen no reason to think that he can carry his burden of proof. Thus, in the following, we shall take it as an a priori truth that the following principle – *Epistemic Constraint* or just *EC* – holds true for moral discourse (about moral principles):

(EC) For all moral *p*, *p* is true if and only if *p* is knowable

So, the prospect of an antirealist – superassertibilist – notion of moral truth is still with us. At least until someone makes sense of the idea of in principle unknowable universal moral principles.

6. *Moral minimalism summarised.*

To sum up, we have so far seen that the defining claims of moral minimalism are the following: Truth is – *pace* the deflationist – a substantial property about which we may conduct meaningful philosophical realism-debate. Moral discourse exhibits the

³² My argument against potentially evidence-transcendent truth here may be a version of Wright's argument from normativity. Wright (1993), p. 23-26.

appropriate syntax and discipline to count as assertoric and, hence, as truth-apt. To be a truth-concept is to comply with certain basic truth-platitudes. Different truth-concepts with different, further, realism-relevant features may satisfy the platitudes in different discourses. But, and this is the claim that defines moral minimalism, moral truth does *not* have any further realism-relevant features. In particular, we have seen that it has the feature that it is a priori that there are no in principle undetectable moral truths – i.e. discourse about (universal) moral facts does not have the realism-relevant feature of verification-transcendence. Under the assumption of this – and noting certain other features of moral discourse – we have also seen that moral truth can be modelled by a concept constructed out of the standards for warranted assertibility that govern moral discourse – a concept of superassertibility.

Now, since moral minimalism involves the claim that moral truth has no realism-relevant features, it also involves the claim that moral truth lacks the Cognitive Command feature. I shall assume that Cognitive Command is the most basic realism-relevant feature a discourse might have in the sense that possession of it is a *necessary* condition for having any *other* realism-relevant feature. It will also become clear that we may not want to say that showing that some discourse *satisfies* this constraint amounts to a defence of a fully fledged *realism* for the discourse. Let me now turn to presenting and motivating the Cognitive Command constraint in more detail and showing (in the remainder of this chapter and in the next chapter) against various critics in the literature that *mere minimalism* – in other words, the very idea of truth without Cognitive Command – with respect to a discourse is a coherent position.

7. *Introducing the Cognitive Command constraint.*

Realist intuition about some area of discourse has it that there is an independent (in the sense of independent-of-human-beings) world of facts ‘out there’ to which opinions held in the discourse aim at corresponding. The idea, in a sense, is that an independently existing reality ultimately disciplines the discourse – enforces the distinction between *seeming* and *being* right. To give some examples, realists will tend to say that discourse about things like the colour and the shape of the objects that immediately surround us are obvious candidates for being such an externally

disciplined discourse. By contrast, the realist will think that offering an opinion formulated within moral, aesthetic or comic discourse is not a question of representing an independent reality.

But, how can we capture this intuitive difference? How do we get a firmer, less intuitive, grip on the idea of an *externally* disciplined discourse? As has already been made clear, the minimalist framework for realism-debates makes room for there being more than one realism-relevant intuition in play in the traditional realism-debates – and therefore allows for there being several different cruxes in the realism-debates. But, what is arguably the *basic* realist intuition is the following: There is only *one* set of facts 'out there' – the world is one – and, thus, everybody offering an ideally competent opinion in a discourse aiming at *representing* these facts ought to *agree* on what they are. At least, that is, if they have a chance of knowing the facts in the first place – that is, if truth in the domain is evidentially constrained. So, substantial correspondence³³ and representationality in a discourse and, in some sense, an expectancy of convergence among suitably placed disputants intuitively go together.

Wright's notion of Cognitive Command is exactly intended to capture this vague intuitive idea of a link between representationality and convergence. This constraint gives as a criterion of (realism-relevant and not merely platitudinous) representationality in a discourse that it is possible to rule out a priori *disagreements* formulated within the discourse that are not down to a cognitive error on the part of at least one disputant. To arrive at a more exact formulation, Wright exploits an analogy with mechanical devices functioning representationally.³⁴ Think, for example, of a camera. It is a platitude that if two purely representational devices diverge in output, then this must be down to divergence in *input* (i.e. that they are not aimed at the same thing) or less than suitable (external) *conditions for functioning* or some sort of *internal malfunction*. In other words, if there is divergence in output, yet neither of these three conditions obtain, then we are not dealing with representational devices after all. In that case, the individual devices somehow *add or subtract* something such that the purpose is no longer to be purely representational.

³³ *Substantial* correspondence, because we are now not just interested in the *platitudinous* sense in which any true proposition (also a merely minimally truth-apt one) 'corresponds to the facts' or 'says things as they are', but in a sense of correspondence that goes beyond mere compliance with the minimal platitudes and accommodates further realist intuitions.

³⁴ See Wright (1992), p. 91-93.

The idea is now that a discourse is genuinely representational if compliance with its standards of warranted assertibility makes the speakers function in a way analogous to such representational devices. If two speakers are applying the standards of warranted assertibility of a genuinely representational (and epistemically constrained) discourse, then, if they have the same evidence and it is gathered under suitable conditions and it is processed correctly, then it is guaranteed a priori that they will represent a given states of affairs in the same way – i.e. that they will agree. In other words, in a genuinely representational (epistemically constrained) discourse, we have an a priori guarantee that there is convergence in opinions under ideal circumstances – that is, circumstances under which the three conditions just given are satisfied, where there are no cognitive shortcomings. It is a priori that if there is disagreement within a discourse of this genuinely representational sort, it must be down to a failure of one of the three mentioned types in one or both disputants.

Wright uses this feature of a genuinely representational discourse when formulating his criterion. A discourse thus exhibits Cognitive Command if and only if:

“It is *a priori* that differences of opinion formulated within the discourse, unless excusable as a result of vagueness in a disputed statement, or in the standards of acceptability, or variation in personal evidence thresholds, so to speak, will involve something which may properly be regarded as a cognitive shortcoming”.³⁵

(I shall comment on the provisos about vagueness shortly). In more detail, we would expect there to be three general kinds of cognitive shortcoming involved in all representational disputes, still thinking in terms of the analogy with representational devices:

- 1) Divergent input – i.e. that the disagreeing speakers are working with different information and, hence, that at least one of them is guilty of error or ignorance.
- 2) Conditions (external) of information-gathering are unsuitable such that one or both speakers are not able – due to distraction and the like – to attend properly to the data and infer correctly from them.
- 3) Malfunction (internal) – such that one or both speakers process the information wrongly, for instance by prejudicial assessment of data.

³⁵ See Wright (1992), p. 144.

So, we have with Wright arrived at one way of cashing out the realist's intuitions about substantial correspondence and representation by exploiting an analogy with representational devices. This calls for some further clarificatory comments. First of all, why is it necessary to look away from various forms of *vagueness*? The answer here will give us a better understanding of the way in which the constraint tests for indeterminacy in a domain. Secondly, why does the criterion test for an *a priori* (as opposed to merely a posteriori) guarantee against cognitively blameless disputes? And, finally, is passing the proposed criterion *sufficient for a fully fledged realism* about a discourse? Let me deal with each in turn.

First, to understand why intractable differences of opinion that are down to semantic or other forms of vagueness should not be allowed to compromise the seriously representational status of a discourse consider that, arguably, *all* discourses – except formalised ones like mathematics – display vagueness of some form (at least semantic vagueness). Thus, discourses that are intuitively seriously representational – for instance, discourse about colour or about medium-sized objects in our immediate surroundings – appear to allow for intractable disagreements without cognitive shortcomings that are down to mere semantic vagueness. So, in order for the Cognitive Command criterion to test for a *realism-relevant* feature of a discourse – and to allow us to make a realism-relevant *distinction* between different truth-apt discourses – we must look away from cognitively blameless disagreements that are merely down to forms of vagueness. Only if we preclude such kinds of intractable disagreement in our formulation of the criterion, will the criterion test for *realism-relevant* indeterminacy in a discourse.

But, what this shows is of course also that recognising that a discourse fails to exhibit Cognitive Command is a way of recognising that there is *indeterminacy* in the discourse which is not down to semantic or other forms of *vagueness*. I shall later in this chapter explore the link between failure of Cognitive Command and indeterminacy.

Secondly, the criterion requires that it be *a priori* that a cognitive shortcoming is involved in any dispute. Why? To answer this question, consider that we can easily imagine having a posteriori evidence that a given discourse is such that as a matter of fact there are no disagreements save ones that involve cognitive shortcomings of the above-mentioned kinds, but where the discourse intuitively is not in the business of

serious representation. Thus, if a virus happened to wipe out everyone except people with one particular sense of humour, then discourse about comedy would as a matter of fact allow only admissible types of disagreement to occur. But, a purely contingent event like this should, of course, not be allowed to decide the status of a discourse in the realism-debates. Therefore, satisfaction of the constraint has to be a conceptual rather than a purely empirical matter. In other words, it has to be true in virtue of the very meanings of the terms from a discourse – and, hence, knowable a priori – that disagreements must be down to cognitive shortcomings in order for the discourse to satisfy a realism-relevant constraint like the Cognitive Command constraint.

An immediate consequence of this feature of the constraint is that claiming that a discourse *fails* to exert Cognitive Command does not, as it has been understood by critics, involve claiming that *there are* (or even that it is *possible* that there are) cognitively blameless disagreements formulable within the discourse. Failing to pass the test is rather a question of our *lacking* the resources to claim that we have *an a priori guarantee* that such cognitive shortcomings must always be present. This observation will play an important role later in this chapter.

Finally, there is a question about exactly what passing the Cognitive Command test amounts to. In what sense is the criterion realism-relevant? In the foregoing it has been argued that genuinely representational (epistemically constrained) discourses are such that it is a priori that opinions formulated within it must display convergence except in cases of cognitive failures. But, strictly speaking, this amounts only to the claim that it is a *necessary* condition for a discourse's being representational that it exhibits Cognitive Command. It is important to notice that there is no argument so far to the effect that a discourse exhibits Cognitive Command *only if* it is genuinely representational. Thus, it is still possible that even *non-representational* discourses – what might be called 'internally disciplined discourses' – can satisfy the constraint.

In other words, what the criterion arguably tests for is *satisfaction of a pre-condition for being* seriously representational rather than for *being* seriously representational. In order for a discourse to actually *be* seriously representational, it is arguably further required that there is a domain of independently existing facts to which the true opinions held within the discourse can correspond. One way of cashing out the notion of independent here would be to apply the Wide Cosmological Role test described earlier. So, what has emerged is that satisfaction of the Cognitive

Command constraint, albeit a necessary condition for realism about a discourse, is not a sufficient condition. One can, for instance, happily be a moral anti-realist *and* a believer in moral Cognitive Command as long as one does not accept what may be called *factualism* about morals where factualism is understood as the claim that moral facts have Wide Cosmological Role.

Thus, there are at least three positions to occupy within the minimalist framework with respect to moral discourse (given that we have already argued against the possibility of an evidentially un-constrained notion of moral truth). 1) One might accept that moral discourse is minimally truth-apt, yet be radically non-objectivist and anti-realist about morals by claiming that moral discourse fails to exhibit Cognitive Command. This is, of course, the minimalist's choice. 2) One might accept that moral discourse is minimally truth-apt and exhibits Cognitive Command, but argue that there are no substantial moral facts – no real correspondence – since the a priori guarantee against cognitively blameless convergence is brought about by a strict *internal* discipline in the discourse rather than by an independently existing, external, domain of facts. I.e. one might be what can be called an *objectivist* antirealist. 3) Or, finally, one might be a genuine moral realist going for genuine representationality in morals. That is, going both for moral objectivism – satisfaction of the Cognitive Command constraint – and moral factualism – satisfaction of the Wide Cosmological Role constraint.

However, the claim that moral minimalism (option 1) is a genuine possibility – that one can consistently believe in moral truth without objectivity and realism – has proved controversial. Since I shall be defending moral minimalism in this thesis, the rest of this chapter will be mainly devoted to a defence of the very stability of this minimalist position.

8. *Cognitive Command and Epistemic Constraint*

Before turning to this defence of the coherence of mere minimalism, however, there are some observations to make about the interaction between the two realism-relevant constraints so far discussed: Cognitive Command and Epistemic Constraint.

That is, with the criterion of Cognitive Command now formulated in some detail, we are in a position to see, first, that discourses in which statements may be

undetectably true or false ought to satisfy the constraint. If we are dealing with a discourse apt for epistemically unconstrained truth, then states of affairs in the world will determine the truth-value for any belief that p regardless of whether we are actually or even in principle able to come to know the truth-value of p . In other words, the world imposes a distinction between seeming right and being right even in cases where we may never as a matter of principle be able to tell the difference. This amounts to an extreme example of externally imposed discipline. If a discourse is in such a way able to represent states of affairs radically beyond our ken, we will certainly expect it to be genuinely representational and so pass the test for Cognitive Command.

But will it – as formulated so far – pass the test? Consider two scientists arriving at contradictory theories. Assume that scientific discourse is apt for undetectable truth and assume also that their two theories incorporate different, and incompatible, views about what counts as an observation – that is, assume that a version of the thesis that all observation is theory-laden is true – such that each scientist can be said to have corroborated their theory by faultlessly applying sound scientific methods. In what *could* the cognitive shortcoming consist here? The only possible answer – given that both scientists arrive at their opinion faultlessly (by their own criteria) – appears to be that the falsity of one or both of the *very beliefs in dispute* will constitute the cognitive shortcoming.

Thus, given that we intuitively want the test to be passed by discourses apt for epistemically unconstrained truth (and granting for the purposes of argument that the thesis about theory-laden observation is true), two things follow for our conception of cognitive shortcoming: Firstly, we have to allow that the fact that the very opinion in dispute is false suffices for ascribing a cognitive shortcoming to its holder. Secondly, we have to accept that the question of *who* is failing cognitively here is itself undecidable – that cognitive shortcomings themselves may be undetectable. Call this *a thin notion of cognitive shortcoming*.

But, what now of discourses in which only an epistemically constrained concept of truth applies, i.e. discourses in which we expect the following principle of epistemic constraint to hold?

EC If 'p' is true then evidence is available (in principle) that p

This is the interesting case in the present context – our investigation of the status of moral discourse – since as has been argued moral truth ought to be regarded as epistemically constrained.

What becomes of the notion of a cognitive shortcoming once we have accepted epistemic constraints on truth? Well, it would seem that we *can* say more than in the case of discourses apt for evidentially unconstrained truth. Say that it is the case that p and that I am ignorant, since I fail to believe that p , or that I am in error, because I falsely believe that not- p . Now, given EC, there will be evidence available in principle for p and that means either – in the case of ignorance – that I have a cognitive shortcoming in virtue of not applying my means of gathering evidence to their full potential or – in the case of error – that I have a cognitive shortcoming in virtue of misapplying these means. Given EC, there is always more to say about the putative cognitive shortcoming involved in a disagreement about p than just that one party has a false opinion. Further cognitive shortcomings relating to the evidence available must be present and they are – in virtue of EC – in principle identifiable. But this a priori expectation surely must be motivated somehow. We must ask for an *explanation* of why we are to expect – a priori – that there will be such an identifiable cognitive shortcoming. In other words, it should be kept in mind in the following that, given EC, the notion of a cognitive shortcoming becomes more substantial than we saw above in the case of discourses apt for undetectable truth – and that the defender of the claim that a discourse exhibits Cognitive Command (given EC) incurs further explanatory duties as a result.

9. *Trivialising the Cognitive Command constraint?*

Critics of the coherence of minimalism claim that satisfaction of the Cognitive Command constraint is guaranteed merely in virtue of a discourse's being truth-apt. Now, a simple thought appears to support this suggestion. If there is a disagreement within a given discourse, is it not guaranteed a priori that at least one disputant has a cognitive shortcoming *simply in virtue of the falsity of at least one of the very beliefs in dispute*? If A believes that p and B believes that not- p , then is there not a guarantee that A or B has a false belief? If this is the case, then of course the constraint can

make no interesting realism-relevant distinction between different truth-apt discourses. Indeed, it will make no distinction at all.

Wright anticipated this trivialising move in *Truth and Objectivity*.³⁶ His original proposal for countering it exploited the before-mentioned feature of the minimalist framework for the realism-debates according to which anti-realism is always the default stance. Assuming this approach, the onus of proof always rests on the realist to demonstrate that truth in a domain has a further realism-relevant feature. So, it is not the job of the defender of mere minimalism about some discourse to show that he can block the above trivialising argument to universal satisfaction of Cognitive Command. Instead, the onus is on the objectivist to provide a detailed argument showing why it is a priori that disagreement within the discourse has to involve a cognitive shortcoming.

If the objectivist carries the burden of proof like this, then establishing that the constraint is satisfied is never a mere triviality, Wright argues. It cannot be justified without carrying out substantial philosophical work. In particular, Wright's claim is that the objectivist cannot succeed in making intuitively non-representational discourses – like moral discourse – pass the Cognitive Command test without defending substantial and *unattractive* philosophical positions. He argues in more detail as follows.

One way for the realist to prove (a priori) that disagreement within a discourse has to involve a cognitive shortcoming is by showing a priori that the discourse is apt for potentially evidence-transcendent truth. As argued in the previous section, if truth is not evidentially constrained, we can ascribe a cognitive shortcoming on the basis that we know that one party or the other holds a false belief – and that without having to identify further cognitive shortcomings with regard to the evidence. But, of course, showing that truth in a discourse is potentially evidence-transcendent is no mere triviality. Anyway, moral discourse, as I have argued above, is unsuited for evidence-transcendent truth.

So, in our present context the question becomes: Can the would-be trivialiser show (a priori) that a cognitive shortcoming trivially must be involved in disputes in any minimally truth-apt discourse, given that it is only apt for evidentially constrained

³⁶ See Wright (1992), p. 148-157.

truth? As argued in the last section, the assumption of EC means that we should expect to be able to give a more *substantial* account of the cognitive shortcoming involved. Ultimately, the trivialiser will have to provide an appropriate epistemology of the discourse in question. And it may be argued – at least for the discourses in dispute between the realist and the anti-realist, including moral discourse – that this will invariably require the provision of an indefensible *intuitional* epistemology (at least if EC holds for the discourse and if the discourse is *sui generis* – irreducible). The following argument for this claim derives from Wright. It will be developed in much more detail in the chapters to come.³⁷

Either the epistemology of the discourse in question is inferential or it is non-inferential (i.e. intuitional). If it is inferential, then either we must be drawing inferences – getting justification – from other beliefs from the same discourse or we must be drawing inferences from beliefs from other discourses. The former cannot be the case indefinitely, however, since that would lead to circularity or vicious regress. So, at some point we must be drawing inferences from beliefs from a different discourse. But, then we will need principles linking these other beliefs with the beliefs in the discourse in question. In order to defend the claim that our original discourse exhibits Cognitive Command, these principles will in turn be required to satisfy the constraint. So, we are back to square one and the whole dialectic will start over again. Thus, in order to stop the regress, we will eventually end up on the other horn of the dilemma – claiming that the epistemology associated with the discourse is intuitional.

But, by being committed to an intuitional epistemology one incurs further – seemingly unbearable – commitments. Wright suggests that we accept the principle that one should not associate intuitional capacities with a discourse, unless the existence of such a faculty is part of our *best explanation* of the fact that people non-collusively agree about the subject-matter of the discourse. Again, it will not be a trivial task to meet this commitment. In particular, it appears to be very difficult in cases where the discourse in question is affectively grounded.

³⁷ The following dialectic works under the assumption that the discourse in question is *disputationally pure*. That is, that there are disagreements within the discourse which are not such that they can be resolved by turning to other discourses. It is – in a few words – the idea that there are disagreements, for instance, about what is funny which turn solely on a sense of humour. And it excludes the possibility that the discourse might be *reducible* to another discourse. (I discuss this notion of disputational purity in much greater detail in chapter 6).

In conclusion, no matter how the realist goes about defending the claim that a discourse exhibits Cognitive Command – and he must defend it, since the minimalist argues that he carries the burden of proof – he will have to do substantial philosophy. It is never a *purely trivial* matter to defend the claim that a truth-apt discourse exhibits Cognitive Command. And, furthermore, the prospects for success in defending evidentially unconstrained truth, intuitional epistemologies or reductive claims for the discourses presently in focus seem dim.

There are potential problems with this strategy for defending the gap between minimal truth-aptitude and Cognitive Command, however. First of all, as mentioned above, it depends on the general approach to realism-debates according to which anti-realism is always the default position. Accordingly it is likely to be rejected as unsuccessful by the many non-subscribers to this principle. One might, for instance, argue that common-sense is a better candidate for the default position. But, of course, that would mean that in the case of moral discourse the onus *would* rest on the realist. Or one might hold, with Shapiro and Taschek,³⁸ that starting out neutral is the proper default when one is trying to develop a *neutral* framework for the debates. However, in the following, I shall argue on the very plausible assumption that the realist carries the burden of proof.

Let me just briefly record a second worry about this defensive strategy. Mark Sainsbury³⁹ raises doubts about whether it is as demanding as Wright seems to think to come up with a workable intuitional epistemology for the discourses at the centre of the debate. Working with a ‘lean’ conception of faculties and claiming that these faculties do not have to be *sui generis*, he holds that there is still hope for an intuitional epistemology for the relevant discourses and, therefore, for the trivialising theorist.

However, a potentially more worrying problem for the defender of the gap between minimal truth and Cognitive Command derives from a more precise formulation of the trivialisation challenge that appears to show that there is a fundamental inconsistency in the very idea of lack of Cognitive Command. Answering this worry will allow us to see more clearly how we are to conceive of the gap. The

³⁸ See Shapiro and Taschek (1996).

³⁹ See Sainsbury (1996). Wright answers in Wright (1996c).

strategy I shall pursue was originally suggested, I believe, by Wright in his answer to an objection made by Timothy Williamson.⁴⁰

10. *Trivialising Cognitive Command: Two arguments.*

In the literature there have been attempts to provide decisive arguments to the effect that all minimally truth-apt discourses must satisfy the Cognitive Command constraint trivially. I shall now consider two arguments for the collapse of the distinction advanced by Williamson and Shapiro/Taschek. Both arguments purport to be reductions to absurdity of the assumption that a discourse fails to exhibit Cognitive Command. Let me present and comment on them in turn.

Timothy Williamson's simple version of the trivialising argument goes as follows.⁴¹ Assume that x 's belief that p is true and assume that y 's belief that $\neg p$ is true. Now, given that any truth-predicate satisfies a kind of Disquotational scheme for beliefs – the principle that a belief that p is true if and only if it is true that p – our assumption leads directly to a contradiction. We must conclude that x 's and y 's beliefs are not both true. But seemingly that will give us reason to conclude that either x or y is wrong about the truth-value of p .⁴² That, in turn, entails – given a thin notion of a cognitive shortcoming – that at least one of them has a cognitive shortcoming. So, it would seem that every minimally truth-apt discourse trivially satisfies the Cognitive Command constraint. Let me formulate a precise version of the argument to work with.⁴³ Call this the First Trivialising Argument:

1	(1) (x believes that $\neg p$ and y believes that p) & $\neg(x$ or y has a cognitive shortcoming).	Prem.
1	(2) x believes that $\neg p$	1, &E
3	(3) p	Ass.
1,3	(4) x is guilty of cognitive shortcoming	2,3 *

⁴⁰ The general strategy I will be pursuing in the following section is anticipated by Wright (1994a). See also his (1994b).

⁴¹ See Williamson (1994).

⁴² Notice that this way of formulating the challenge makes it depend crucially on the step from ' x and y are not both right about p ' to 'Either x is wrong or y is wrong about p ' and this step uses classical logic. The De Morgan step from $\neg(p \& q)$ to $\neg p \vee \neg q$ uses DNE and in the following, I will – to anticipate – argue that the validity of classical logic is *not* to be assumed in the present context.

⁴³ This version of the argument is anticipated in Wright (1994a), 333-34. Shapiro/Taschek (1996) run a similar argument. It is not clear to me that their argument adds anything to Wright's improvement of Williamson's argument, though Wright seems to think so, see Wright (1996), 929n.

1,3	(5) \perp	1,4, \neg E
1	(6) \neg p	3,5, \neg I
1	(7) y believes that p	1,&E
1	(8) y is guilty of cognitive shortcoming	6,7 *
1	(9) \perp	1,8, \neg E
	(10) \neg [(x believes that \neg p and y believes that p) & \neg (x or y has a cognitive shortcoming)]	1,9, \neg I

There are at least three observations to make about this argument.

First of all, the immediate focus of attention should be the steps from (3) to (4) and from (7) to (8). These steps clearly presuppose what was dubbed the thin notion of a cognitive shortcoming – the idea that someone’s merely having a false belief (i.e. the very belief in dispute) could be sufficient for ascribing to him a cognitive shortcoming. But in doing this, the steps also seem to presuppose that no matter what discourse p is taken from (including moral discourse), the faculties involved in forming beliefs about the subject-matter of that discourse are purely cognitive.

Faced with this argument and these presuppositions one might feel compelled to respond on behalf of defender of the gap between minimal truth-aptitude and Cognitive Command by attempting to find an independent way of determining whether the abilities exercised in a discourse are cognitive or not. That is, if it has been established independently that some of the abilities associated with for example discourse about comedy are non-cognitive, then someone's merely having a false belief about what is funny will not suffice for ascribing a *cognitive* shortcoming to him. For example, it could be that beliefs about the comic are grounded in a *non-cognitive* sense of humour. That is the type of reply which Williamson may have thought his trivialising argument demanded.

But, if this is the way to defend the gap, things look gloomy for the defender. It is very difficult to see how such an independent account of what a cognitive ability is can be given without begging the question of whether the discourse in question deals in ‘real matters of fact’ or not. We will simply be left with the problem we started out with – i.e. to find a principled way of distinguishing between genuinely cognitive discourses or abilities and the rest. So, the attempt to block the trivialising argument by objecting to the steps from (3) to (4) and (7) to (8) by developing independent criteria for an ability's being cognitive seems doomed from the outset.

A second comment to make is that line (10) strictly speaking is not a statement of satisfaction of Cognitive Command. The former has the form of a negated disjunction whereas the latter is a conditional. Thus the challenger must take the argument further than the conclusion in line (10) to reach a straightforward statement of satisfaction of Cognitive Command. In particular, in order to get a proper statement of Cognitive Command as the conclusion of the argument, the trivaliser will have to take the following steps to line (18). (The essential move here is of course the one from ‘ $\neg(p \ \& \ \neg q)$ ’ to ‘if p then q’):

	(10) $\neg[(x \text{ sincerely believes that } \neg p \text{ and } y \text{ sincerely believes that } p)$ $\ \& \ \neg(x \text{ or } y \text{ has a cognitive shortcoming})]$	1,9, \neg I
11	(11) x believes that $\neg p$ and y believes that p	Ass.
12	(12) $\neg(x \text{ or } y \text{ has a cognitive shortcoming})$	Ass.
11,12	(13) (x believes that $\neg p$ and y believes that p) $\&$ $\ \neg(x \text{ or } y \text{ has a cognitive shortcoming})$	11,12 $\&$ I
11,12	(14) \perp	10,13, \neg E
11	(15) $\neg\neg(x \text{ or } y \text{ has a cognitive shortcoming})$	12,14 \neg I
11	(16) x or y has a cognitive shortcoming	15, DNE
	(17) (x believes that $\neg p$ and y believes that p) \rightarrow $\ (x \text{ or } y \text{ has a cognitive shortcoming})$	16, \rightarrow I
	(18) AP [x believes that $\neg p$ and y believes that p \rightarrow $\ \text{either } x \text{ or } y \text{ has a cognitive shortcoming}]$	17, Apriorisation

Of course, line (18) amounts to a statement of the satisfaction of Cognitive Command. (The last step (to 18) is justified by noting that line (17) rests on no premises. Hence, line (17) is provable and can be arrived at by means of logical and conceptual reflection alone – in other words, it is knowable a priori).

Now, what this demonstrates is that the first trivialising argument – in order to threaten the truth-aptness/Cognitive Command gap – *must rely on classical logic*. In particular, the step from line (15) to (16) uses DNE. I shall later argue that this assumption of classical logic is controversial in exactly the discourses about which the realism-antirealism controversy is most alive.

A third comment to make about the first trivialising argument is that it fails to take into account the fact that the discourses that invite an antirealist reading – in particular candidates for failure of Cognitive Command – are discourses for which truth is arguably epistemically constrained. This has the immediate consequence – as already noticed – that there is no need to work with a thin notion of a cognitive shortcoming. However, a second trivialising argument – presented in the literature by

Shapiro and Taschek – attempts to show that Epistemic Constraint on truth coupled with failure of Cognitive Command also leads to inconsistency. Let me turn to this version of the trivialisation challenge.

Shapiro and Taschek advances the following argument – we can call it the Second Trivialising Argument.⁴⁴

1	(1) For any p, if p then it is knowable that p	EC
2	(2) (x sincerely believes that $\neg p$ and y sincerely believes that p) & $\neg(x$ or y has a cognitive shortcoming).	Failure of CC (?)
3	(3) p	Ass.
1,3	(4) It is knowable that p	1,3, $\rightarrow E$
1,2,3	(5) x has come to believe something that is not only false but the negation of something knowable	2,4
1,2,3	(6) x is guilty of a cognitive shortcoming	5
1,2,3	(7) \perp	2,6, $\neg E$
1,2	(8) $\neg p$	3,8, $\neg I$
1,2	(9) It is knowable that $\neg p$	1,8, $\rightarrow E$
1,2	(10) y has come to believe something that is not only false but the negation of something knowable	2,9
1,2	(11) y is guilty of a cognitive shortcoming	10
1,2	(12) \perp	2,11, $\neg E$
	(13) EC is inconsistent with failure of Cognitive Command	1,2,12

This argument also calls for further comment. As with the first trivialising argument, it does not so far establish that Cognitive Command holds for every truth-apt discourse. It only establishes that it is not the case that line (2) – a formulation of definite failure of Cognitive Command – is true. But a direct statement that Cognitive Command holds true universally can only – by a line of reasoning identical to the one rehearsed above in connection with the first trivialising argument – be reached under the assumption of classical logic. So, in this respect the two arguments lead to the same conclusion, though the second argument goes through without relying on a thin notion of a cognitive shortcoming.

With the two trivialising arguments against the possibility of a gap between minimal truth-aptitude and satisfaction of Cognitive Command now laid out before us in a precise form, we can see more clearly what the commitments of a defender of the gap must be. That is, he has to deal with two problems:

First of all, he must give reasons for disqualifying classical logic – opting for intuitionistic logic instead – for the discourses about which he wants to be a

⁴⁴ Shapiro and Taschek (1996), p. 85. I have modified the wording slightly.

minimalist. If not, as we have seen, it is possible to prove that a discourse satisfies the constraint in so far as it is truth-apt.

Secondly, however, this will not suffice as a defence of the gap. What we have seen above is also that *the very idea of a cognitively blameless disagreement appears to be inconsistent*. Both arguments demonstrate that the assumption of a cognitively blameless disagreement (i.e. line (1) in the first argument and line (2) in the second) can be reduced to absurdity (i.e. line (9) in the first argument and line (12) in the second). In other words, the defender of the gap must not claim that failure of Cognitive Command is constituted by the possibility (or actuality) of a definite cognitively blameless disagreement. The challenge to the defender is therefore to characterise failure of Cognitive Command in some alternative way. The remainder of this chapter is devoted to showing that the minimalist can deal with these two problems.

11. Meeting the trivialising challenges: Agnosticism about classical logic.

Now, the best minimalist answer to these challenges to the very idea of a merely minimally truth-apt discourse is to focus on the fact that the Cognitive Command constraint strictly speaking states that for discourses satisfying the constraint *it is a priori* that any dispute formulated within the discourse involves a cognitive shortcoming on the part of one or both participants in the dispute. Thus, strictly speaking *failure* to meet the constraint is constituted by there failing to be (so far at least) an a priori case for such cognitive determinacy. To claim that a discourse fails to satisfy the constraint is to profess *agnosticism about the apriority* of the cognitive determinacy within the discourse – not, as the above trivialising arguments seem to suggest, to claim that there are or could be disputes that are cognitively blameless. Let me now show how this observation helps the minimalist to live up to his commitments as outlined above.

Let me start out by arguing that – given this observation – the minimalist can avoid holding that it is a priori that classical logic is valid in the disputed contexts. The minimalist can base this agnostic stance towards the validity of classical logic on the following two claims: 1) We must *remain agnostic* about the apriority of the claim that moral discourse is complete – i.e. we must remain agnostic about having an a

priori warrant for the claim that, for any moral proposition p , evidence will, at least in the best epistemic circumstances, decisively tell for or against p . That is, we must remain agnostic about the following claim (where $\Diamond Kp$ means that p is knowable, i.e. that evidence is (in principle) available for p):

(Completeness) It is a priori that $\forall p (\Diamond Kp \vee \Diamond K\neg p)$

To argue for this claim, notice that if someone remains agnostic about whether it is a priori that all disputes must be cognitively blameworthy, then he will also have to remain agnostic about the apriority of the claim that the discourse is complete (or decidable). Therefore, claiming that completeness holds a priori for the discourse is to beg the question against the minimalist. The right attitude – if one is undecided about whether it is a priori that any dispute within a discourse must be cognitively blameworthy – is to remain agnostic about the completeness of that discourse. 2) The second claim is the already argued point that moral truth is epistemically constrained – that is, satisfies EC:

(EC) 'p' is true if and only if it is knowable that p

But, if we accept these claims, I shall now argue, we can have no a priori guarantee of the validity classical logic (in particular, LEM, DNE and bivalence) in the moral context. We must be logical revisionists.

The argument for revisionism goes as follows.⁴⁵ First of all, contraposing on the aforementioned principle EC, we get:

(EC₋) If it is not knowable that p then it is not the case that 'p' is true

The platitudes about truth – essentially the Disquotational Schema – give us the Negation Equivalence:⁴⁶

(NE) 'It is not the case that p' is true if and only if it is not the case that 'p' is true

Now, combining these principles we get:

⁴⁵ The argument can be found in Wright (1992), p. 41-44.

⁴⁶ As was argued in section one above.

(a) If it is not knowable that p then 'it is not the case that p' is true

And, by applying EC on the consequent of (a):

(b) If it is not knowable that p then it is knowable that 'it is not the case that p'

But then the following line of reasoning can be initiated:

1	(1)	$\neg\Diamond Kp \rightarrow \Diamond K\neg p$	(b) above, from NE and EC
2	(2)	$\neg\Diamond Kp \vee \Diamond Kp$	an instance of LEM
3	(3)	$\neg\Diamond Kp$	Ass.
1,3	(4)	$\Diamond K\neg p$	1,3 \rightarrow E
1,3	(5)	$\Diamond K\neg p \vee \Diamond Kp$	4 \vee I
6	(6)	$\Diamond Kp$	Ass.
6	(7)	$\Diamond K\neg p \vee \Diamond Kp$	6 \vee I
1	(8)	$\Diamond K\neg p \vee \Diamond Kp$	2,3,5,6,7 \vee E

But line (8) amounts to a proof of completeness – hence the claim that it is a priori that all matters in the domain are decidable under ideal circumstances. And that contradicts our first assumption above – that we are dealing with a discourse where we precisely lack an *a priori* guarantee that all matters are thus decidable.

To avoid the contradiction – while holding on to the involved principles: the platitudinous Negation-Equivalence, EC and agnosticism about the *a priori* status of the principle of completeness – we are compelled to change our logic. That is, the only plausible way of blocking the problematic move to (8) – given independent arguments for EC and agnosticism about the *a priori* warrantability of completeness – is now an intuitionistic refusal to assert LEM. That is, to remain agnostic about the a priori status of LEM. In short, the minimalist can exploit the fact that the following three theses can be held consistently:

- (1) Epistemic constraint on moral truth.
- (2) Agnosticism about the a priori status of LEM in moral discourse.
- (3) Agnosticism about the a priori status of completeness of moral discourse.

And thus he is not compelled to affirm the second consistent triad:

- (1) Epistemic constraint on moral truth.
- (2) It is a priori that LEM holds true in moral discourse.
- (3) It is a priori status that moral discourse is complete.

But, then we have an answer to the first challenge to the minimalist position – a reason to disqualify the relevant parts of classical logic (e.g. LEM, DNE) in the context of moral discourse. To assume a priori the validity of classical logic – in the light of EC – is to beg the question against the moral minimalist. It is to assume that completeness holds true a priori and, hence, also to assume that it is a priori that any disagreement must involve a cognitive shortcoming – the very thing at issue!

12. Meeting the trivialising challenge: Characterising indeterminacy.

The minimalist defender of the gap between truth-aptness and Cognitive Command – employing now the strategy based on the observation that, on a careful reading of the Cognitive Command constraint, it states that it is *a priori* that disagreements in the discourse involve a cognitive shortcoming – is not out of the woods yet, though. He still has to deal with the fact that any claim about determinate *counterexamples* to the principle that a truth-apt discourse satisfies the Cognitive Command constraint leads to a contradiction – as both trivialising arguments showed. Why exactly – one might now ask – should he not want to accept that failure of Cognitive Command amounts to the claim that there are *determinate counterexamples* to the truth of the constraint?

The issue here is an instance of the more general issue about how best to characterise *indeterminacy* – into which I shall now shortly digress. That is, the claim that a discourse fails to exhibit Cognitive Command is one way of claiming that the discourse allows for indeterminacy. It is precisely the suggestion that a discourse may be indeterminate in the sense that we must remain agnostic about the claim that it is a priori that disagreements formulated within the discourse involve a cognitive shortcoming. But, before settling on this characterisation of indeterminacy, let me look at what alternative ways there are for characterising indeterminacy. I shall briefly consider two alternative characterisations of indeterminacy that I ultimately will want to avoid – truth-value-gap theories and epistemicist theories.⁴⁷

On the first alternative approach to characterising indeterminacy, it is claimed that indeterminacy implies a truth-value-gap. I.e. that an indeterminate *p* is *neither*

⁴⁷ This is also the strategy of Wright (1994b).

true nor false and therefore also – given the standard disquotational schema – that there are counterexamples to the Law of Excluded Middle. The claim – implicit in the two trivialising arguments above – that there are determinate counterexamples to satisfaction of the Cognitive Command constraint amounts, in the light of EC, to claims that there are such truth-value-gaps.

The gappy theorist, however, has two immediate problems to contend with. First of all, assuming the existence of a gap – assuming that, for some p , $\neg(p \vee \neg p)$ – results in absurdity in both classical and intuitionistic logic:

1	(1)	$\neg(p \vee \neg p)$	Prem.
2	(2)	p	Ass.
2	(3)	$p \vee \neg p$	2, $\vee I$
1,2	(4)	\perp	1,3 $\neg E$
1	(5)	$\neg p$	2,4 $\neg I$
1	(6)	$p \vee \neg p$	5, $\vee I$
1	(7)	\perp	1,6 $\neg E$

(Again, this is, of course, the same inconsistency as the one that troubles the characterisation of failure of Cognitive Command as the existence or possibility of definite counterexamples). Now, there are (non-intuitionist) ways of being revisionistic about logic that avoid this problem of inconsistency for the gappy theorist.⁴⁸ But, at least there is a *prima facie* problem for the gappy theorist – since we should not be revisionist if there are alternative and satisfactory *conservative* ways of

⁴⁸ In particular, a logic that works with a weakened *reductio ad absurdum* principle can avoid the inconsistency. Thus, the *reductio* principle ('T' stands for 'it is true that'):

$$\begin{array}{l} \text{(WR = Weak} \\ \text{Reductio)} \end{array} \qquad \frac{X, A \vdash \perp}{X \vdash \neg TA}$$

blocks the move to the contradiction in the following way:

1	(1)	$\neg(p \vee \neg p)$	Prem.
2	(2)	p	Ass.
2	(3)	$p \vee \neg p$	2, $\vee I$
1,2	(4)	\perp	1,3, $\neg E$
1	(5)	$\neg Tp$	2,4, WR
6	(6)	$\neg p$	Ass.
6	(7)	$p \vee \neg p$	6, $\vee I$
1,6	(8)	\perp	1,7 $\neg E$
1	(9)	$\neg T\neg p$	6,8 WR
1	(10)	$\neg Tp \ \& \ \neg T\neg p$	5,9, $\&I$

Of course, the bottom line no longer amounts to an inconsistency! Thanks to Patrick Greenough for pointing this strategy out to me.

characterising indeterminacy – that should motivate us to look elsewhere for a characterisation.⁴⁹

Secondly, there is a more general problem with the gappy theory of indeterminacy. What the gappist invites us to think is that there are three statuses that a proposition may have: determinately true, determinately false and determinately neither true nor false. The question is now whether we want indeterminate cases to have a determinate status like this – whether this is not to miss the very phenomenon of indeterminacy. An indeterminate case should rather be a case where *it is not determinate* whether it is true or false. Determinate truth-value-gaps simply fail to capture the phenomenon. So, again an alternative approach is best sought.

On the second alternative approach to indeterminacy, it is claimed that indeterminacy is merely epistemic – that p or not- p is always determinately true, but that we may in some cases never be able to determine which. This claim involves rejecting EC for the disputed discourse. Williamson is himself an example of someone who takes such an epistemicist, determinist, approach in the case of semantic vagueness. All questions about the truth-value of vague statements do according to him have determinate answers. Seemingly indeterminate cases are always determinately true or false. It is just that they may be so beyond our ken. Vague discourse would, of course, also on this view exhibit Cognitive Command, since as we saw earlier evidence-transcendent truth ensures satisfaction of the constraint. (In the light of this, Williamson will not feel compelled by the argument for disqualifying classical logic in the context of vague discourse). Notice now, however, that an analogue of this move – the move of making facts about meaning evidence-transcendent and thus saving classical logic in the face of incompleteness – will not be available when thinking about moral discourse, if what I have argued independently about the epistemic constraints on ethical truth holds good.⁵⁰ So, again, we should look elsewhere for a way of characterising moral indeterminacy.

⁴⁹ As it will become clearer in the following chapters, allowing for truth-value-gaps in *moral* discourse is a *particularly* un-attractive position. The fact that moral judgements guide action and that a moral consensus is of great practical importance for the agents – for reasons of social co-operation and harmony – it is often not an option to settle for there being no answer to some moral question. We tend – I claim – to always go on to try to find a common solution to a moral question, rather than just agree to differ.

⁵⁰ For Williamson's view consult Williamson (1994 b). See also Wright (1994b) for more criticism of Williamson's approach to semantic vagueness. A further criticism of epistemicism is that the epistemicist's characterisation of vague discourse is ultimately committed to a very strong and, I

So, thus far, there are good negative reasons to look to the minimalist's third way of characterising moral indeterminacy and to claim merely that, about indeterminate *p*, we must remain agnostic about an *a priori guarantee that all disagreements involve a cognitive shortcoming*. And we can now see clearly why the minimalist should reject the characterisation of failure of Cognitive Command that the two trivialising arguments implicitly assumed that he endorsed – the characterisation that involved claiming the existence of definite counterexamples. This way of characterising failure of Cognitive Command ultimately fails to capture the phenomenon of moral indeterminacy, since it implies the thesis that there are *determinate truth-value-gaps* which, for the independent reasons given, is an unhappy way to deal with the phenomenon.

13. Conceptual knowledge and modality: A final objection?

So, the position that is to be defended for the remainder of this thesis is that moral truth is merely minimal truth – where this involves remaining *agnostic about the apriority* of the principle that all moral disagreements are cognitively blameworthy. So far, it has been argued against the trivialising arguments that this is a consistent and non-trivial position. The more positive argument for the claim that this is the *best* way to capture moral indeterminacy – and the way to formulate a viable form of cognitivist anti-realism – will emerge in the coming chapters of this thesis.

However, let me before going on consider a possible objection to this third proposal that will clarify the position considerably. The objection asks whether the minimalist's claim that *it is not a priori* that all disagreements are cognitively blameworthy (and someone remaining agnostic about the apriority of this principle will allow that this is at least a possibility) does not entail the claim that it is *not necessarily* the case – or equivalently that it is *possibly not* the case – that all disputes are cognitively blameworthy. If either of these claims *are* entailed by the minimalist's position, he will be in the same trouble as those who hold that failure of Cognitive Command amounts to there being definite counter-examples.

believe, indefensible version of externalism in the philosophy of language. But, I will not take on that issue here.

To see this in a simple way, let me draw a parallel. The dialectical situation encountered by the defender of the gap has parallels to the one faced by the intuitionists in their attempts to resist endorsing the Law of Excluded Middle. As we have already seen above, an outright denial of LEM – i.e. the claim that there are strong counterexamples to LEM – is not a consistent way of formulating the intuitionist's worry. Consider now also that it does not help the intuitionist to go modal – for instance, to merely assert *the possibility* of the negation of LEM – i.e. $\Diamond\neg(p \vee \neg p)$, the possibility of strong counterexamples:

1	(1)	$\Diamond\neg(p \vee \neg p)$	Prem.
2	(2)	$\neg(p \vee \neg p)$	Ass.
3	(3)	p	Ass.
3	(4)	$p \vee \neg p$	3, $\vee I$
2,3	(5)	\perp	2,4 $\neg E$
2	(6)	$\neg p$	3,5 $\neg I$
2	(7)	$p \vee \neg p$	6, $\vee I$
2	(8)	\perp	2,7 $\neg E$
1	(9)	\perp	1,2,8 $\Diamond E$

Notice here also, that it will not help the intuitionist to claim that ‘it is not necessary that $(p \vee \neg p)$ ’. This is equivalent – given the rule for equivalence between the modal operators – to ‘it is not not possible that not- $(p \vee \text{not-}p)$ ’ and that can also be seen to generate contradiction for the intuitionist just by adding an extra step to the above derivation:

$$(10) \quad \neg\Diamond\neg(p \vee \neg p) \qquad 1,9 \text{-I}$$

So, the intuitionist has to be cautious and avoid any direct denial of LEM. There are no coherent strong counterexamples to LEM – not even modalised ones. And, of course, a parallel point can be made about modalised versions of statements about cognitively blameless disputes. Failure of Cognitive Command cannot coherently be expressed by a forthright denial that a cognitive shortcoming is involved in any case of disagreement – that is, by giving definite counterexamples – *nor by any statement about the possibility of such counterexamples*.

So, the minimalist will have to argue that his agnostic attitude about the a priori status of the statement that all disputes involve a cognitive shortcoming does

not entail any claim about the modal status of the statement. In particular, the following principle must not be valid:

(M₁) It is not a priori that $p \rightarrow$ It is possible that not p

Which – given that 'it is possible that not' is equivalent to 'it is not necessary that' – is to say that the following contraposed principle must be invalid too:

(M₂) It is necessary that $p \rightarrow$ It is a priori that p

Fortunately, for the minimalist, this principle is already challenged in the literature on good grounds. That is, given that the modality involved above is *metaphysical* modality, there is no longer the presumption that what is necessarily the case – what is true in all possible worlds – also must be knowable without recourse to empirical knowledge about the actual world, i.e. also must be knowable a priori. This is, of course, a lesson that many have drawn from the writings of Kripke and Putnam.

Let me give an example that will make it clearer why we should not accept these principles – in particular (M₁).⁵¹ We do not as a matter of fact have a proof of – that is, have *a priori* warrant for – for example the following claim:

(G) The extensions of the predicates 'is a counterexample to Goldbach's conjecture' and 'is a prime-number between 32 and 36' coincide.

This is because we do not at present have a proof that Goldbach's conjecture has *no* counterexamples – i.e. that the predicate 'is a counterexample to Goldbach's conjecture' applies to no set of two numbers. But from this lack of an *a priori* warrant for (G) we cannot infer that it is possible that the two predicates diverge in their extensions. In our present state of ignorance about the truth or falsity of Goldbach's conjecture, the claim that it is a possibility that the two predicates diverge in extension is unjustified, since, if the statement is true, it will be *necessarily* true. So, when we are agnostic about the a priori status of some statement, it does not automatically follow that it is possible that it is false – i.e. (M₁) is false. And therefore there is no danger that the minimalist's agnosticism about the a priori status of the claim that all moral disputes are cognitively blameworthy will collapse into the claim

⁵¹ Also to be found in Kripke (1972), p. 36-38 and Wright (1993), p. 460.

that it is possible that there are cognitively blameless moral disputes – and hence inconsistency. The final possible objection has been answered.⁵²

So, we have now identified, in the abstract at least, a position in the debates about realism and objectivity – that of the mere minimalist – and we have seen that it involves no inconsistency, *pace* certain objectors in the literature. The claim to be defended in the following is, thus, that moral discourse is adequately and coherently described both as truth-apt and as failing to exhibit Cognitive Command. Thus, it is implicitly argued that the interesting realism-relevant contrast to draw in meta-ethics is that between discourses exhibiting Cognitive Command and *merely* minimally truth-apt discourses – i.e. the contrast between discourses about which it is *a priori* that no cognitively blameless disagreement can occur and discourses about which we remain agnostic about the claim that it is *a priori* that no cognitively blameless disagreement can occur. This, I am alleging, is the best way to formulate an anti-realist, but cognitivist position about morals. In the following chapters the position will become more developed and, I hope, come to seem overall plausible. Thus, it will be argued that it can deal with, is consistent with, important features of moral discourse – the alleged possibility of radical moral disagreement and moral dilemmas, the tension between moral ontology and naturalistic accounts of moral discourse, the supervenience of the moral on the natural. In the end, I hope to have developed and motivated a new and viable form of ethical anti-realism – to compete with the

⁵² The strategy considered here for defending the notion of Cognitive Command against the trivialisation threat essentially hinges on the assumption that questions about whether a discourse satisfies the constraint or not should be answered *a priori* – by means of a conceptual analysis of the core concepts of the discourse. Williamson’s rejoinder attacks the very idea of *a priori* answers to this type of questions. He asks: Can Wright’s disputes about realism – in particular, about Cognitive Command and the Euthyphro Contrast – be decided without sometimes taking into account purely empirical states of affairs? By way of example he suggests that the empirical discovery of the relativity of colour to our sense-apparatus told us something about the realism-relevant properties of discourse about colour that pure conceptual analysis – *a priori* investigations – could not have done. See Williamson (1996). I shall not go into this dispute here. Only notice – with Wright (1996c) – that if empirical-scientific discourse is going to decide matters about the status of other discourses in the realism-debates, then the status of empirical-scientific discourse itself should be determined in advance as objective. But, how is this to be determined, if not by conceptual means only? I will assume in the following that philosophical questions about the objectivity-status of discourses are *conceptual* questions.

problem-ridden alternatives of expressivism and error-theory. First, however, there is a second challenge to the just proposed minimalist framework to be answered.

Chapter Two

Minimalism, Expressivism and Motivation

Moral minimalism, as we have now seen, is a form of *cognitivist* anti-realism about moral discourse. The cognitivism of the position clearly renders it at odds with a rival formulation of moral anti-realism – the expressivist or non-cognitivist position. In fact, the minimalist framework just presented makes non-cognitivist-style anti-realism a non-starter in the debates. However, an argument has been presented that tries to establish that non-cognitivist moral anti-realism ought still to be considered a live option by the minimalist. This argument is thereby also a threat to the very consistency of the framework adopted in this thesis. A defence of the proposed minimalist framework for the realism-debates against this argument is therefore necessary before I can explore the *mere* minimalist position formulated within it.

Let me briefly outline the argument that appears to present a threat. The distinctive claim of expressivist forms of moral anti-realism is that moral sentences do not have genuine assertoric content. Thus, to sincerely utter a sentence from ethical discourse is not, according to the non-cognitivist, to express a belief. Rather, it is to *express* a non-cognitive state – a desire, an emotion or the acceptance of a norm – or, in other versions, it is to *prescribe* a certain way of acting. Thus, the views of philosophers such as the emotivist Ayer, the prescriptivist Hare, the quasi-realist Blackburn and the norm-expressivist Gibbard fall under the general heading of ethical non-cognitivism or expressivism.⁵³

As mentioned, from the point of view of the framework adopted in this thesis these positions have become obsolete. The adoption of a minimalist approach to truth and truth-aptness entails the rejection of the distinctive claim of moral non-cognitivists. That is, the minimalist argues that all assertoric discourses are apt for (at least minimal) truth. And, since the minimalist's minimal requirements for having assertoric content – namely that a discourse is disciplined and displays appropriate syntax – are met by the sentences of ethical discourse, we must conclude that the non-

⁵³ See e.g. Blackburn (1993), Gibbard (1990), Hare (1981). (Though Hare 'gets extremely cross' when he is classified as a moral non-cognitivist. See Hare (1999), p. 4. And perhaps it is unfair to

cognitivists got it wrong. Moral discourse *is* assertoric and apt for truth and, hence, moral sentences *are* apt for expressing beliefs. Thus, if we change the framework of the realism-debates in accordance with the minimalist's suggestions – by introducing a minimalist conception of truth and truth-aptness – expressivism becomes a non-starter in the debate. Minimalism about truth gives us cognitivism on the cheap, it would seem.⁵⁴

What some philosophers have asked, however, is whether it can really be the case that a well-entrenched and widely held position in meta-ethics can be dismissed in such an easy manner – and in a way that does not even engage in substantial meta-ethical debate. In this chapter, I will be concerned with attempts by Frank Jackson, Graham Oppy and Michael Smith – the Australians, for short – to block this fast track repudiation of expressivism.⁵⁵ In a series of articles, they have tried to exploit the fact that the minimalist is committed to a notion of belief via the notion of assertion. Their crucial point is that it is possible for the expressivist to hold that, since the states expressed by moral sentences *cannot be beliefs* due to their internal connection with motivation and action, ethical sentences cannot have assertoric content despite the fact that they are disciplined and have the appropriate syntax. It is in this sense that, ultimately, they present an argument against the whole minimalist framework that is being applied in this thesis. And it is as such that it requires a reply on behalf of the minimalist.

In the first section, I try to sharpen the Australian objection to the minimalist's proposed rejection of expressivism. Their objection turns out to be based on what I shall call the standard argument for expressivism (section 1). Assessing the arguments of the Australians involves taking a closer look at two background assumptions in the expressivist's argument. This accounts for the two following sections – on the Humean theory of motivation and moral motivational internalism and on a variety of ways of resisting the standard argument for expressivism (sections 2 and 3). With the help of the distinctions drawn in these sections, I then attempt to defend a distinction between merely minimal belief and more robust belief, since this, in turn, is

classify him as a non-cognitivist, given that he also appears to applaud Wright's minimalist account of truth. See Hare (1999), p. 16).

⁵⁴ This argument for the exclusion of the expressivist position from the realism debates as they are conceived of by the minimalist is given by Wright himself. See e.g. Wright (1992), p. 35-37.

⁵⁵ See e.g. Jackson, Oppy, Smith (1994), Smith (1994a) and (1994b).

the distinction required in the defence of the minimalist's position (section 4). With this distinction in place, it is possible to block the Australian argument without engaging in substantial philosophy – and thus retaining the fast track to a repudiation of expressivism (section 5). It is finally suggested that all this gives the minimalist a dialectical advantage over the expressivist – given that the latter faces the famous Frege-Geach-Hale problem.⁵⁶

1. Minimalism and expressivism: The argument for compatibility

Let me, first, try to get the argument put forward by the Australians into clear focus – the argument purporting to show that expressivism *is* compatible with minimalism about truth and truth-aptness after all. The Australians claim that Wright has not realised that one of the platitudes in the network surrounding the concept of truth, namely the analytical link between assertoric status or truth-aptness and aptness for expressing *belief*, will cause trouble for his account of truth-aptness. In particular for the claim that *moral* discourse is truth-apt.

Assertoric or truth-apt sentences are apt for expressing beliefs. That is, beliefs have as their content a representation of the way things are and representations of the way things are, are exactly what assertions or truth-apt sentences are apt to give expression to. Therefore, the minimalist had better concede as analytically true that if somebody sincerely asserts that *p*, then this person believes that *p*, or to put it in terms of truth-aptness:

(P) If *s* is minimally truth-apt, then *s* can be used to express the content of a belief

Wright recognises that this is a platitude. But, according to the Australians, this leaves the door open for the expressivist in a way overlooked by Wright. What (P) states is that being apt to give the content of a belief is a necessary condition for being truth-apt. But, what I shall call *the standard argument for expressivism* has exactly the conclusion that this necessary condition is not met in the case of the states expressed by moral utterances. A version of this standard argument goes:

⁵⁶ For a classical discussion see Hale (1986). See also the exchange between Hale and Blackburn in Haldane and Wright (1993) and Wright (1995).

- (1) The states expressed by the sincere utterance of moral sentences are intrinsically motivational (when about the obligations of the subject himself)
 - (2) No *belief* is intrinsically motivational.
-

∴ The states expressed by the sincere utterance of moral sentences are not beliefs.

(By ‘intrinsically motivational’ here is meant something like ‘motivational in its own right’ in the way desires as opposed to beliefs are often thought to be – more on the Humean claim that there is such an asymmetry between belief and desire later). Let me briefly introduce the two premises of the expressivist's argument. The first premise amounts to the assertion of a strong version of moral motivational internalism – the claim that moral judgements are in some intimate way (to be discussed in more detail later) connected with motivation. Suffice it for now to give an example. According to the motivational internalist, somebody who sincerely professes to believe that he ought not to lie, but also shows in his behaviour that he does not intend to avoid lying (even in situations where there are no other overriding moral concerns in play), is unintelligible or irrational. The second premise is a corollary of what I (with Michael Smith) shall call a Humean theory of motivation which is constituted by the two theses:

- a) Only desires are intrinsically motivational.
- b) Beliefs and desires are distinct entities.

Let me now try to give a preliminary assessment of the dialectical situation between the minimalist and the Australians. According to the Australians, the expressivist can reappear in the dialectical landscape by showing that appropriate syntax and discipline is not sufficient for a discourse's being truth-apt as long as there are independent grounds for claiming that the states expressed by the utterance of sentence from the discourse are not beliefs. And, such independent grounds are available in the case of moral discourse – it is argued – in the shape of the premises of the above argument. In other words, it is only by being *too* minimal about truth-aptness – i.e. by not properly respecting the analytical principle (P) and its consequences – that the minimalist can hold that minimalism immediately excludes expressivism.

The challenge posed to the minimalist who wants to exclude expressivism, then, appears to be to show why the expressivist is not allowed to run the above argument to the conclusion that moral discourse is not truth-apt. On the face of it, the minimalist can do this only if he denies the expressivist one or both of the premises in his argument – that is, if he rejects the strong version of motivational internalism or rejects in one way or another Humeanism about motivation. But, then it would seem that his original claim – that minimalism about truth automatically excludes expressivism – will have to be qualified. We must say that minimalism about truth and truth-aptitude excludes expressivism *only if strong motivational internalism or Humeanism about motivation is false*. So, to establish that moral sentences are truth-apt, the minimalist will have to commit himself to one or another substantial theory in the philosophy of moral psychology. As Smith formulates it:

‘Without such a philosophy of mind, the minimalist does not have a theory that tells us which sentences are truth-apt at all; he is like a cook who has a recipe but no ingredients.’⁵⁷

But, in that case, there is a sense in which minimalism is not minimal:

‘...minimalism about truth in fact *presupposes* an understanding of folk psychology, an understanding of a theory whose concepts of belief and desire will in turn constrain which sentences can and cannot count as truth-assessable. But if minimalism presupposes an understanding of folk psychology this way then, given that that theory is a substantive theory about the way human beings work, it follows that minimalism itself requires substantive assumptions in giving accounts of truth and truth-assessability’.⁵⁸

What should the minimalist say to this? The straightforward option is to stick to his original account of truth-aptness – according to which discipline and appropriate syntax are the only requirements for truth-aptness – and then tailor the notion of belief after this. That is, he can accept the analytical connection between assertion and belief, the principle (P), by committing himself to a class of beliefs – call them *minimal* beliefs – identifiable solely by studying the syntax and discipline of the sentences used to express them.⁵⁹ This strategy for answering the Australians is hinted

⁵⁷ Smith (1994b), p. 24.

⁵⁸ Smith (1994b), p. 24.

⁵⁹ Wright suggests that the minimalist might adopt Blackburn’s term ‘commitment’ here. Wright (1998a), p. 190.

at by Wright and Paul Horwich and there have been attempts to develop it in some detail by John Divers and Alex Miller.⁶⁰ Thus, Wright writes in a response to Jackson that the proper answer to the objection above is to say that "...*any* attitude is a belief which may be expressed by the sincere endorsement of a sentence which complies with the constraints of syntax and discipline...".⁶¹ Then the minimalist's strategy would be to say that sentences from moral discourse are apt for expressing minimal beliefs and that all the expressivist can hope to show is that they are not apt for expressing beliefs in *a more robust sense of belief* – where robustness of belief requires robustness of the truth-predicate that governs the discourse in which it is expressed.

I believe that this strategy of introducing a notion of minimal belief is essentially right. The aim of my discussion is, therefore, to argue that *the introduction of a notion of minimal belief will answer the Australian objection on behalf of the minimalist*. This will involve responding to the Australians' suggestion that either this concept of minimal belief is a mere decoy-concept or it is implicitly tied to substantial theoretical commitments in the philosophy of mind. Before that, however, I will turn my attention to the premises of the standard argument for expressivism. How are the two premises supposed to work together and what are the options for somebody who wants to resist the conclusion?

2. *Humeanism about motivating and normative reasons.*

Some introductory remarks about the second premise – the Humean theory of action – are needed before I go on to discuss how the minimalist can answer the Australian. I believe it is possible to isolate at least three independent theses which have all been defended as part of a Humean theory. I will discuss the third thesis – concerning the Humean account of *normative* reasons – later. But I start with the two theses that we have already met under the name of a Humean theory of motivation – or of *motivating* reasons to use Smith's term.⁶² They are:

⁶⁰ Wright (1994c), Divers and Miller (1994) and (1995), Horwich (1994).

⁶¹ Wright (1994c), p. 170.

⁶² Let me briefly introduce Smith's distinction between motivating and normative reasons here, since I shall be using it in the following. First, we must distinguish the two different perspectives from which one can give an explanation of an action – the *intentional* and the *deliberative* perspective. From the *intentional* perspective we give *teleological explanations* of actions. We can say with Smith that we account for the agents actions by citing the motivating reasons behind them – by

- i) Only desires are intrinsically motivational.
- ii) Belief and desire are distinct existences.

Of course, the exact content of these theses depends on how we define 'desire' and 'belief'. And this is not a trivial matter. We may thus, first of all, distinguish a *narrow* and a *wide* conception of 'desire'. And, importantly, we must notice that the Humean theory of motivation above concerns desires *in the wide sense*.

In the wide sense, it is sufficient for an attitude's being a desire to ϕ that it is a 'pro-attitude' towards ϕ -ing – i.e. that it is in some way something that moves us to ϕ . Michael Smith suggests another – roughly equivalent – way of characterising this wider concept of desire, namely as 'an attitude that the world must fit rather than an attitude that must fit the world'. The idea is here that any propositional attitude is a desire if it is such that, if there is a discrepancy between the propositional content of the attitude and the world, then it is the world that is to be changed and not the attitude. Both these wider definitions will allow us to say that attitudes deriving from a sense of duty or loyalty, for instance, are desires-in-the-wide-sense though these attitudes often animate us to do unpleasant things that we do not desire-in-a-narrower-sense to do. Also, it allows us to say that *non-occurrent* attitudes are desires. Attitudes that we do not attend to consciously most of the time – for instance a continuing desire to survive – are still pro-attitudes or attitudes that the world must fit rather than the opposite.

Often, however, 'desire' is defined more narrowly and the Humean theses above become more controversial as a result. One way of narrowing the notion of a desire is by favouring an account of desire according to which desires necessarily have phenomenological content – such that something can be a desire only if it is *felt* and therefore also is occurrent. One attempt to develop such a narrower notion of desire

referring to the psychological states (beliefs and desires) which caused the action. From the *deliberative* perspective, on the other hand, we try to *justify* actions through practical reasoning. Thus we consider a set of truths – abstract and not psychologically real entities – which together form a practical syllogism. Its premises will be of the general form 'x ought to ϕ ' and 'x can ϕ by ψ -ing' and the conclusion will be of the general form 'x ought to ψ ' – or in Smith's terminology: 'x has normative reason to ψ '. These truths (true propositions) can, of course, in turn be the objects of our belief-attitudes and we then have normative beliefs. Thus we can say that going through a process of deliberation can result in our coming to believe that somebody has a normative reason to do something – that a certain action is rationally justifiable or desirable or valuable for someone.

is found in Scanlon (though he finds little use for it in his own account).⁶³ He speaks of an agent's 'having a desire in the directed-attention sense that P' if the thought that P *keeps occurring to the agent* in a favourable light. Clearly, only a subclass of states that the world must fit keep occurring to agents in a favourable light in this way. Arguably, this narrower notion is closer to the intuitive or everyday notion of desire than the wide notion employed above. Clearly, it excludes non-occurrent, longer-term desires. And it excludes desires to ϕ out of a sense of duty, for instance, where ϕ -ing does not occur to the agent in a favourable light – or may even seem downright unpleasant to the agent. For example one may feel one has a duty to convey an unwelcome truth to a friend and strongly desire-in-the-narrow-sense not to.

The relation between desire and motivation can, however, be conceived of weakly or strongly in a further way – independently of the question of whether desires are conceived of as necessarily attention-directing/phenomenal or not. According to a strong, Humean, conception, beliefs alone can never cause us to act, since motivational force *ultimately resides* in desires (widely conceived) – they are *always* the ultimate source of the motivation. According to a weaker conception, it is merely the case that desires are *always present* whenever there is motivation. This latter reading allows that the desires are merely necessarily *consequentially ascribed* in cases where it is beliefs that ultimately motivate.

To understand the strong conception here, it is not enough to notice that there always has to be a desire somewhere in the causal background of an action. Desires alone cannot be sufficient for action either. If we do not know how to achieve what we desire – i.e. have an appropriate means-end belief – we cannot be motivated to act. So, a complete motivating state – a state sufficient for action – consists in a belief/desire-pair. But, and this is the point of the first Humean thesis *read strongly*, it can be argued that there is an *asymmetry* between beliefs and desires with respect to the role they play in producing action.

Thus, what the strong Humean holds is that it is always the desire that ultimately does the motivating. The claim is that desires are motivating states in their own right, whereas beliefs are only extrinsically motivating states – the motivating force of a belief is contingent on the presence of a relevant (intrinsically motivating)

⁶³ Scanlon (1998), p. 39.

desire. Perhaps drawing an analogy to a car with its accelerator and clutch is helpful here to convey the Humean picture. Having a desire without having the appropriate means-end belief is, according to the Humean, like stepping on the accelerator of a car while keeping the clutch down. Adding a means-end belief is then to be likened with letting go of the clutch. The latter merely enables the engine to engage with the wheels, i.e. the actions of the agent, but does not add any power to the engine. All the power – the ability to move the car – resides in the accelerator and engine, that is, in the desires.⁶⁴

Thus, on the strong, Humean, version of the first thesis, only desires can be ultimately motivational. By contrast, if the first thesis is conceived of weakly, we can accept that a desire may be merely consequentially ascribed. I have a consequentially ascribed desire to ϕ , if the desire is ascribed merely because I am cognitively motivated. In that case, that is, the motivation is ultimately a consequence of having a set of beliefs – the beliefs being the ultimate *source* of the motivation. Later, I shall discuss McDowell's and Dancy's theories in more detail according to which desires are in this way merely consequentially ascribed.

Looking back at the standard argument for expressivism, it is clear that the thesis that helps to underpin the second premise of the argument is the first thesis of the Humean theory of motivation *read strongly* (with desire read widely). The weak conception of the relation between desire and motivation, on the other hand – the reading that merely states that a desire is present whenever there is motivation – appears to be a mere platitude, as I shall go on to argue later.

Let me comment, briefly, on the second thesis of the Humean theory of motivation – the claim that belief and desire are distinct existences – that also underpins the second premise of the expressivist's argument. The Humean holds that it will always be possible to pull belief and desire apart modally, i.e. to point to a possible world in which a person has the belief and not the related desire. Hence, the second thesis rejects any idea of belief and desire being necessarily connected or of

⁶⁴ I should add one further minor qualification here. Reason – our beliefs – *can* help us develop desires in one sense, according to the Humean. Given a desire – say my desire to quench my thirst – my means-end-reasoning may cause me to acquire certain further so-called *derived* desires – say the desire to buy a coke. But it is part of the point of the first thesis that we cannot have an infinite regress of derived desires. So strictly speaking, there will always have to be at least one *underived*, intrinsically motivating, desire – a desire not arrived at through means-end-reasoning – present when an action is performed.

there existing unitary states – sometimes called ‘besires’ – which can function simultaneously as belief and desire. A *besire* can, that is, be defined as a psychological state which is belief-like with respect to one content and desire-like with respect to another content. To give an example, it is ruled out by the Humean that there can be a necessary connection or an identity between x's belief that *p* – that it is wrong to lie – and x's desire that *q* – that lying is avoided.

The third thesis that may be part of a Humean account of action is the Humean account of *normative reasons*.⁶⁵ A short formulation of the third thesis could be:

- iii) An agent has a (normative) reason to ϕ only if he has a *natural* desire to ϕ .

In other words, according to this Humean thesis, one can *justify* ϕ -ing only if one has a natural desire to ϕ . A *natural* desire can be defined as a desire that exists, not as the result of any rational activity of the agent, but simply as a product of the brute workings of our nature. Consequently, the contrasting concept is that of a *rationally acquired or a motivated* desire – a desire created (or removed) by a process of, moral or prudential, reasoning.⁶⁶ Thus, moral and prudential reasoning, according to this Humean idea, is constrained by what desires we are already endowed with by nature (though, as already mentioned, there may be *derived* desires involved). All justification for actions must be grounded in certain natural or brute – even irrational – desires that are beyond the reach of rational criticism. This Humean thesis also goes under the name of the 'maximizing account' of normative reasons, since it involves the claim that our reasoning faculties can only help us to achieve maximum satisfaction of our already existing natural desires.

3. Three ways of resisting the standard argument for expressivism.

On the face of it there are at least three different ways of resisting the standard argument for expressivism. That is, it is possible to reject its different component

⁶⁵ Smith (1994c), p. 94-98.

⁶⁶ This is not to say that the Humean believes that our un-derived natural desires remain constant through our lives. They may change through our encountering new things in the world or through natural developments of our personality, perhaps. But such change cannot according to the Humean be brought about rationally, say by realising through moral deliberation that one ought not ϕ .

theses – the claim that moral attitudes are motivational and the Humean claims that only desires are intrinsically motivational and that beliefs and desires are distinct. I now turn to an overview and a discussion of these possibilities – with examples of philosophers who have chosen these different options.

Let me first consider the first Humean thesis – i.e. i) above about the relation between desire and motivation. Assuming that desire is to be understood widely here, there are, as mentioned, two versions to consider – the weak and the strong. So first, how might we reject i) when construed weakly? This thesis is the claim that I will never be moved to act, if I do not also desire (widely) that the world will turn out one way rather than another. Consider what Smith calls a 'powerful' three step argument *for* this thesis:

- (a) Having a motivating reason is, *inter alia*, having a goal.
- (b) Having a goal is to be in a state with which the world must fit.
- (c) Being in a state with which the world must fit is desiring

As Smith remarks, (a) appears to be a conceptual truth and (b) follows from the fact that realising that a goal has not been reached does not make one give up this goal (via the definition of states with which the world must fit). (c) is perhaps more controversial on the face of it. That is, if one holds that there are states with both directions of fit – in the manner of a moderate anti-Humean who rejects the second Humean thesis and, therefore, believes in besires (as defined in the previous section) – one would claim that being in a state with which the world must fit is *either* to be desiring *or* to be besiring. However, this does not essentially challenge the first Humean thesis weakly read since, as moderate anti-Humeans will concede, when somebody has a besire it is always possible also to ascribe to him a desire, since a besire is a belief that *is* a desire. So even if one challenges (c) above on account of being a believer in besires, one will not therefore have challenged the first Humean thesis read weakly. The upshot of this argument, I shall later conclude, is that the thesis is a platitude, provided that desire is understood widely and the thesis is read weakly.

One option, however, for someone who still wants to reject this weak version of the first thesis is to claim that the wide conception of desire makes the concept 'utterly vacuous' and that this concept must therefore be rejected. Mark Platts has

attempted to do just this.⁶⁷ In that case, the first Humean thesis, of course, would become highly controversial and non-platitudinous. There are, that is, certainly cases where one wants to ϕ without desiring to ϕ in the narrow or phenomenological sense – think again of the example of being duty-bound to convey an unwelcome truth to a friend despite desiring (narrowly) not to. But, why does Platts think that the wide conception has to be vacuous? It seems to me at least to be useful to define ‘having a desire’ as being in a state (with propositional content) that the world has to fit, as opposed to the opposite. For one thing, it helps us to formulate the first Humean thesis such that it becomes a platitude. It allows us to see what is platitudinous in the Humean theory. How *could* Platts possibly argue in a way that forbids us to define a concept of desire in this way?

Notice also, that other seemingly *radical* anti-Humeans – someone like Dancy⁶⁸ – will find Smith's argument for the first Humean thesis weakly construed ultimately sound. Dancy proposes to fundamentally redesign our description of actions. According to Dancy's radically anti-Humean theory, what motivates is always two representations – two cognitive states. One is a representation of the way the world is now, the other is a representation of 'the world as it will be when and if the action is successfully completed' – in other words, a representation of 'before' and a representation of 'after'. Then he goes on to say that 'what motivates in the case of action is always the gap between the two representations [beliefs], and the occurrence of the desire is *the agent's being motivated* by that gap'.⁶⁹ The desire is thus not on this account a separate active entity in the causal background of an action. It is not what is doing the motivating, but merely present as consequentially ascribed.

How does this square with Smith's argument? What Smith's argument establishes is exactly what the first Humean thesis weakly construed comes down to – namely:

(d) If x has a motivating reason to ϕ then x has a desire to ϕ

⁶⁷ Platts (1979), p. 256.

⁶⁸ Dancy (1993), see also Price (1988).

⁶⁹ Dancy (1993), p. 19.

Smith does not establish that desires *are* (identical to) motivating reasons. All he can get from the argument is that there must be a desire present whenever there is motivation. And that is compatible with Dancy's theory. Thus, Dancy says things like: '...it is hard to see that we should find no role for desire at all in our account of motivation'.⁷⁰ So even a very radical anti-Humean finds reason to believe that having a desire is in some sense a necessary condition for being motivated – it is just that he does not accept that the desire is what ultimately is doing the motivating as Humeans think.

But, notice also that, therefore, Dancy's theory is incompatible with the first Humean thesis *strongly construed*. According to Dancy, desires are not the ultimate movers in the story about an action as the Humean will have it. Rather, they are merely necessarily ascribed as a consequence of there existing cognitive motivation. So, Dancy is an example of someone who challenges the first Humean thesis strongly construed.⁷¹

Let me turn now to the second Humean thesis and ask how this might be challenged. As we shall soon realise, this discussion is intimately related to the discussion of how to formulate the internalist intuition that allegedly supports the first premise of the standard argument for expressivism. Consequently, the discussions of the second and the third ways of resisting the standard argument for expressivism are not clearly separated in the following. I shall take McDowell as an example of someone who rejects the second Humean thesis (and, therefore, also the third thesis). And I shall contrast his theory with Michael Smith's theory – a theory that affirms the first two Humean theses and rejects only the third Humean thesis.

Let me try to characterise these two different anti-Humean positions in a little more detail. The positions share at least one assumption. That is, they are both (motivational) cognitivists in the sense that they hold that beliefs can be sufficient for action. The *difference* between them lies in the way they fit the desires into the picture. McDowell appears to deny that the desire is a distinct existence. That is, the desire is somehow identical with a motivating belief – the belief is thought to be a desire. And, to complete the picture, since it is possible to acquire desires rationally,

⁷⁰ Dancy (1993), p. 19.

⁷¹ McDowell (1979) and (1981) is also someone who affirms the first Humean thesis weakly construed and rejects it strongly construed. I shall discuss his views shortly.

then there will be reasons to act that are not dependent on the existence of natural desires – hence McDowell’s theory also contradicts the third Humean thesis.

The less radical anti-Humean, Smith, on the other hand, accepts the Humean tenet that desires and beliefs are distinct existences. Smith believes that the desires involved in moral motivation do not in a sense pull their own weight in it. Rather, Smith’s picture is that a moral belief contingently causes us to have an independently existing desire which in turn leads to action. What makes this theory anti-Humean is the implicit rejection of the third Humean thesis. That is, according to this picture it is possible for desires to be *created* (and destroyed) as a consequence of the acquiring of normative beliefs – thus there are reasons that do not derive from natural desires.

The difference between these two kinds of anti-Humeanism – McDowell’s and Smith’s – will become clearer, as now I turn my attention to the first, internalist, premise in the standard argument for expressivism. The difference between the positions of Smith and McDowell derives exactly from different ways of interpreting our intuitions about the conceptual connection between moral beliefs and motivation. Their differences, as we shall see, turn on how one should make room for the phenomenon of practical irrationality – various kinds of incontinence or weakness of the will – in one’s account of the internalist intuition.

McDowell is making a very *strong* internalist claim when he claims that moral attitudes are intrinsically motivational. According to him, we cannot even be said to have a clear grasp of moral concepts without also being in a state sufficient for action in relevant circumstances. To support this strong version of internalism McDowell introduces the Aristotelian character called the virtuous man. The virtuous man’s purely cognitive moral conception will *necessarily* cause him to act whenever the relevant circumstances obtain. There is no way he can fail to act on his moral conviction, since he is defined as somebody in whom moral motivating reasons ‘silence’ all other motivating reasons and thus as somebody who will never allow his moral judgement and motivation to be clouded by irrelevant desires. In contrast, somebody who turns out not to be motivated to act according to his (all things considered) moral belief must simply have lost a clear grasp of the ethical concepts involved in his judgement. His moral vision must have been ‘clouded by desire’ and he must have lost the ability, he used to have, to *see* what is right or wrong in the

particular case. If the desire is missing, the belief must be missing too – desire and belief are joined with necessity.

The question is, however, whether this Aristotelian character can help McDowell in his defence of the strong internalist claim implicit in his account. That is, the claim that:

Necessarily: If x judges that it is right for him to ϕ then x is motivated to ϕ .

To see why this version of internalism is difficult to defend, we should consider more closely the gap between one's having a belief about a normative reason and one's performing the recommended action. *If* there is such a gap – if the agent acts against his better judgement – the agent is irrational. Davidson has named this phenomenon weakness of the will or incontinence.⁷² In short, it is doing something intentionally while knowing that one is free to do something different that is *all things considered* the right thing to do. I shall argue that there are two distinct phenomena of interest here. There is the case where one has a normative reason to act in certain way, but where the motivation is lacking – where the agent is suffering from various kinds of depressions or spiritual maladies that sap the motivation, to use Stocker's terms.⁷³ And there is the case where one has a motivating reason – a desire – to act in a way that one cannot justify rationally – as in the phenomenon of overwhelming desires. It is the first of these that poses a problem for McDowell.

The first case is the situation where a belief about what is all things considered desirable fails to issue in a motivating reason to act in accordance with the belief. According to McDowell's strong version of internalism above, this is an impossible scenario. However, arguably, people in certain states of mind *do actually* temporarily lose the ability to be moved by normative beliefs that usually motivate them. These states are different forms of spiritual or physical maladies like *accidie* (which is perhaps best translated as moral weakness), depression, despair, timidity, lack of confidence, extreme tiredness, physical illness or drunkenness. These conditions appear to be able to sap an agent's desire to do what he still genuinely believes he ought to do. Stocker describes the phenomenon thus: 'People suffering from these

⁷² Davidson (1980b).

⁷³ Stocker (1990), p. 225-26.

conditions all too naturally set their sights lower than they need....They naturally turn away from what is best, as being too much or too good for them, or simply as not being for them'.⁷⁴ Smith describes the state of mind of somebody suffering from this condition: '...they may stare the facts that used to move them square in the face, appreciate them in all their glory, and yet still not be moved by them'.⁷⁵ It is plausible to suggest with both Smith and Dancy that this is a fact of moral phenomenology. But, McDowell's strong version of internalism flies in the face of this phenomenological fact and so strong internalism must be deemed phenomenologically unacceptable.

To get this objection to McDowell into clearer focus, I should distinguish the phenomena of *accidie*, depression and so on from the other, already mentioned, form of practical irrationality – namely the case of overwhelming desires or *akrasia*. This is the case where an (all things considered) normative reason *does* issue in the corresponding desire, but where this desire is irrationally overwhelmed by other desires – due to some form of psychological compulsion, physical addiction or emotional disturbance. Perhaps we can most easily make sense of this phenomenon by recalling the notion of desires narrowly conceived – desires that are attention-directing. If some desires keep presenting themselves to us in a favourable light, it is conceivable that, though we may be clearly aware that they are rationally overridden by other, stronger but less insistent, reasons, we end up acting on them exactly because they are so insistent.

Now, McDowell may believe that he *does* have an answer to this latter phenomenon of overwhelming desires – an answer based on the Aristotelian notion of a virtuous man. A virtuous person, that is, will not allow the motivating reason based in his all things considered normative reason to be overwhelmed (though of course he does accept that a normative moral reason can be rationally overridden by an even stronger normative moral reason). In fact, in a virtuous person, a moral motivating reason does not even have to struggle with other desires, since he is exactly the person whose moral motivating reason (based on an *all things considered* normative belief) simply *silences* all other motivating reasons. So, there is no possibility for the virtuous person to fail to act in accordance with his conclusive moral belief. However,

⁷⁴ Stocker (1990), p. 225.

I believe that it is doubtful that this notion of a virtuous man can ultimately help McDowell, though I will not discuss the question in detail here. Suffice it for now to point to the objection that the notion involves a massive idealisation and the question obviously poses itself why we should be interested in how ideal persons – of a type that is never actual – are motivated and act as opposed to how *real* persons are motivated and act.

In conclusion, there is an argument to suggest that McDowell's strong version of internalism fails because it fails to make room for the phenomenon of *accidie*, depression and the like – where this phenomenon is understood as distinct from the phenomenon of overwhelming desires. To generalise this point, these phenomena will pose a problem to anyone like McDowell who wants to make the connection between moral belief and motivation necessary – any believer in desires, in particular. So, it seems that we have not found a plausible way to challenge the second Humean thesis about the distinctness of beliefs and desires.

The previous discussion, however, allows us to formulate the internalist intuition properly. What the phenomena of *accidie*, depression and the like show us is that we sometimes simply fail to desire to do what we genuinely believe that we ought to do. Thus, we must accept a weaker formulation of motivational internalism. This formulation of internalism is arguably platitudinous and it is, thus, a priori that:

Necessarily: If x judges that it is right for her to φ then x is motivated to φ or
 x is practically irrational.

I turn, finally, to Michael Smith who *does* want to accommodate the phenomenon of *accidie* (and the like) by accepting the weaker interpretation of the internalist intuition.⁷⁶ As we shall see, he shows that this, platitudinous, weak internalism can be reconciled with moral cognitivism and an acceptance of a Humean theory of motivation – the first and second Humean theses – if we reject the Humean account of normative moral reasons.

To understand Smith's weak anti-Humean position, we need to take a closer look at the interplay between the two perspectives from which one can evaluate and

⁷⁵ Smith (1994c), p. 123.

⁷⁶ The following is based on Smith (1994c), chapters 4 and 5.

explain an action. One and the same action may be *explained* by citing a motivating reason and *justified* by providing a normative reason for it – by giving a practical argument for it. But, what can we say about the relation between what we rationally decide we ought to do through a process of normative reasoning – our normative beliefs about normative reasons – and what we are motivated to do – our motivating reasons?

Smith takes the internalist intuition weakly interpreted to show that moral beliefs *other things being equal* cause us to have motivated desires. The *ceteris paribus* clause allows him to avoid the pitfalls of McDowell's account by allowing him to avoid the claim that the link between moral belief and motivation is necessary. Smith's central idea is – to put it in other words – that our beliefs about what we have normative reason to do (what we can justify through practical reasoning) *contingently* cause us to have corresponding motivating reasons. More precisely the causal link holds exactly to the extent that we are rational – i.e. do not suffer from phenomena like *accidie*.

The question now becomes how Smith accounts in detail for this contingent causal link while respecting that the connection between moral belief and desire is, of course, conceptual (*qua* platitudinous) as opposed to merely empirical. More precisely, he has to show how to get from 1 to 3 below:

- (1) x believes that he has a normative reason to ϕ (i.e. that he ought to ϕ)
- (2) x believes that he would desire to ϕ if he were fully rational
- (3) x desires to ϕ (or x has a motivating reason to ϕ)

Let me give rough sketch of how he attempts to do this. (1) is the type of belief we acquire when we engage successfully in practical deliberation. The link between (1) and (2) is given platitudinously according to Smith and this seems quite plausible. The move from (2) to (3), however, is evidently not a straightforward entailment – that is what the phenomena of *accidie*, depression and the like show us. But what Smith argues we *can* say is:

($\forall x$) (If *x is fully rational* and if x believes that he would desire to ϕ
if he were fully rational then x desires to ϕ)

This principle is platitudinous, Smith claims, and therefore he can hold that there will be a link between (2) and (3) *in so far as the agent is fully rational* – in particular in so far as he does not suffer from *accidie* or moral weakness. The acquiring of normative beliefs cause us to acquire the related desires exactly to the extent that we can call the person in question rational. Therefore, there is a conceptual link between having a moral belief and being motivated accordingly.

Smith's position is, thus, the most weakly anti-Humean position available. His account squares with a Humean theory of motivation. He holds that we must have an underived – though sometimes rationally acquired – desire as the ultimate motivator in order to be motivated to act. And he does not postulate the existence of *besires*. The relation between moral beliefs and moral desires is according to him a contingent causal relation holding between *distinct* entities: x's belief that he ought to ϕ and x's desire to ϕ . In short, he affirms the first two Humean theses (even when the first is read strongly) and shows that they can be squared with the weak, platitudinous, moral motivational internalism. Smith's anti-Humeanism resides solely in his claim that normative reasoning is not restricted to the attempt to maximise satisfaction of natural desires. We may in our deliberations have to challenge any *prima facie* underived desire and, thus, we *can* create new underived desires through rational deliberation and we *can* get rid of old underived desires. In other words, he rejects the third Humean thesis – the Humean account of normative reasons.

In conclusion to this section, we can now distinguish the first three ways of resisting the standard expressivist argument. The first possibility is to reject the argument by negating its first premise (strong internalism) while affirming the second premise (the corollary of the two Humean theses of motivation). Thus, as we have just seen, Smith has shown how one can reconcile cognitivism and Humeanism about motivation by being a weak internalist. But, we have also seen two ways of resisting the argument by rejecting the second premise namely by rejecting one and/or the other of the Humean theses that support it. We saw Dancy who denied that desires are what ultimately motivate moral actions. And we saw McDowell who agreed with this, but in addition argued that beliefs are not always distinct from desires.

However, *all* these ways of resisting the standard argument for expressivism as part of a defence of the minimalist position have a serious flaw according to the

Australians. If any one of these ways of dealing with the objection is chosen, the minimalist will have to grant that his rejection of expressivism does not proceed directly from his adoption of the minimalist framework for the debate, but rather from taking up substantial positions in the debate about moral psychology. The framework he adopts is hence – contrary to his intentions – no longer neutral. It would seem that an alternative way of answering the Australians is wanted. I shall now show that there *is* such an alternative way with the objection – based on the adoption of the before-mentioned distinction between *minimal* and *robust* beliefs. The answer to be given in the final two sections of this chapter is, to anticipate, that the notion of belief that figures in the second premise of the standard argument for expressivism – and consequently the premise itself – can be seen to be ambiguous once this distinction has been defended.

4. Minimalism and minimal belief

I left the issue of whether minimalism can properly exclude expressivism after the minimalist had just introduced the notion of minimal belief. And ‘minimal belief’ was defined in such a way that any attitude that answers to publicly acknowledged rightness-conditions and is expressed in discourse with an appropriate assertoric syntax qualifies as a minimal belief. The question is whether this notion will help the minimalist to defend the framework that excludes the expressivist from the debate *without further non-minimal commitments*.

In order to answer this question, the minimalist will first of all have to defend the viability of the very notion of a minimal belief. Jackson, Smith and Oppy argue plausibly that in order to do this the minimalist must give an account of how his concept of minimal belief can sustain all the platitudes concerning the *psychological* function of belief. They suggest the following list of platitudes:

- i) Beliefs are aimed at fitting the way things are.
- ii) Beliefs evolve rationally under the impact of information.
- iii) Beliefs can only produce action if they combine with desires.

The Australians now pose a dilemma for the minimalist. If the minimalist cannot accommodate these platitudes in his treatment of the notion of minimal belief, he will

have failed to give us reason to call the states expressed by minimal assertions *beliefs*. In that case, the Australians claim "...[m]inimal belief is no more belief than decoy ducks are ducks".⁷⁷ But, on the other horn of the dilemma, if he *does* manage to coin a notion of minimal belief which sustains all the platitudes about belief, then – according to the Australians – he must have made substantial metaphysical and theoretical commitments and, thus, can only exclude the expressivist from the debate by being willing to defend these commitments and thereby fail to be neutral.

I will take it as my first aim to show that the minimal conception of belief *can* sustain the psychological platitudes – following the strategy of choosing the second horn of the dilemma proposed by Divers and Miller with minor modifications⁷⁸ – while also arguing that following this strategy does not mean that the minimalist incurs substantial commitments. The greatest challenge in showing how the concept of minimal belief can be platitude-respecting is no doubt to show how it can accommodate the third platitude about the relation between belief and desire. But let me briefly say something about how the minimalist could deal with the first two platitudes.

The strategy used in dealing with the first platitude is analogous to the one employed by Wright when trying to show how a minimal concept of truth can accommodate what is genuinely platitudinous about truth as correspondence. Hence, just as there is a perfectly harmless, metaphysically light-weight sense in which saying that a sentence (or proposition) is true is equivalent to saying that it corresponds to the facts, there will be a perfectly harmless and metaphysically light-weight sense in which any true minimal *belief* can be said to correspond to the facts. Observe, first, that it is analytical that:

Things are as represented by the proposition that p if and only if p

(This is of course the analogue of the principle that "Things are as 'p' says they are if and only if 'p' is true"). Likewise we have the platitudinous equivalence – call it the Equivalence Schema for beliefs:

The belief that p is true if and only if p

⁷⁷ Jackson, Oppy, Smith (1994).

⁷⁸ Divers and Miller (1994) and (1995).

This immediately gives us:

The belief that p is true if and only if things are
as represented by the proposition that p

But, then, can we not allow ourselves to paraphrase this as the Correspondence Principle for Beliefs? After all, the relation of correspondence is just the flip-side of the relation of representation. Thus we have:

CP_B The belief that p is true if and only if the belief that p
corresponds to (or fits) the facts

In other words, it seems that the innocent way in which the notion of minimal truth can sustain the correspondence platitude transfers smoothly to the notion of minimal belief.

Do minimal beliefs tend to evolve rationally under the impact of information – as they should according to the second platitude? The minimalist would appear to have the resources to save this intuition in virtue of the fact that any assertoric discourse is disciplined. The participants in a disciplined discourse will be required to assert only what the standards of warranted assertibility allow them to assert given their current informational state. And if their informational state alters, their warrant will do the same. Since our beliefs are intimately linked to dispositions to utter sentences that express these beliefs, it would seem that this constraint on what one can warrantably assert is also a constraint on what one can warrantably believe. Now, to be rational is partly, one might simply argue, to let oneself be restricted by such standards of warrant for belief and assertion. If the statements of a discourse are subject to such a constraint of rationality, it does seem plausible that the beliefs they are apt to express are subject to the same constraint. In other words, there is at least an intuitive case for claiming that the assertoric discipline possessed by any discourse apt for minimal truth shows that minimal beliefs will tend to evolve rationally.

But, as already mentioned, the real problem for the minimalist appears to be to give an account of how his notion of belief can sustain the platitudes about the relation of minimal belief to desire. Jackson, Oppy and Smith give us two versions of the desire-belief platitude. According to them, it is analytical that beliefs "combine

with desires in order to guide us around the world" and that "they guide us in realising our desires". The question is what exactly is platitudinous about the relation between beliefs and desires. My strategy will be to reject iii) as a platitude and then to show that the minimalist can accommodate what is *genuinely* platitudinous about the relation between normative belief and desire.

As Divers and Miller notice, what appears to be implicit in Jackson, Oppy and Smith's proposal is something like the following principle:

(3a) A belief can guide us around the world only if it is accompanied by a desire

This formulation brings out that the proposal is in fact a version of the Humean theory of motivation and normative reasons. The metaphors of beliefs 'combining with desires' and 'guiding desires' do indeed suggest a very strong version of Humeanism – that is, it is suggested that the desires are there totally independent of any beliefs and that the beliefs *only* can play an instrumental role in the production of action. This amounts to *affirming all the three Humean theses*, I distinguished earlier. But, it is unlikely that a controversial philosophical theory like the Humean has the status of a conceptual truth – for one thing that would implausibly make most of the theories of motivation and justification on the market counterintuitive (including Smith's own).

Divers and Miller – who also reject (3a) as simply a version of Humeanism – suggest the following as being the relevant platitude instead:

(3b) It is part of our concept of belief that beliefs *pretty often* have to combine with desires in order to guide us around the world.

This suggested platitude does *not* predetermine the outcome of the debates in favour of the Humean. In fact, it seems to favour an anti-Humean account. However, arguably, (3b) does not have *the generality* of a genuine conceptual truth.

What *I* want now to call attention to is the point made already in the previous section. Namely that there is a platitudinous weakened version of the first Humean thesis – the thesis that desires *always* have to be present when somebody acts. This platitude has the generality that (3b) misses and is clearly not as strong as (3a). In order to argue for this proposal, let me briefly sum up what emerged in the previous section about the different theories about the relation between belief and desire in the

production of action – i.e. theories of motivation. At least four different accounts were distinguished.

(a) There was the suggestion of the defender of the Humean theory of motivation and normative reasons that beliefs are purely instrumental and that they need to combine with independently existing, natural desires in order to produce both reasons and action. Then, there was (b) the proposal that normative beliefs can be sufficient for action (that is, without combining with natural desires), but only because they contingently cause an independently existing – ultimately motivating – desire in a rational agent which in turn causes him to act. This was the 'motivated desire' stance taken by Smith by rejecting the third Humean thesis. (c) And there was the 'besire-theorist' who believed that normative beliefs can be sufficient for action, since they are peculiar belief-states-that-are-also-desires capable of causing action by themselves (guided by means-end beliefs, of course) – thereby rejecting the second Humean thesis about the distinctness of desire and belief as well as the first Humean thesis. (d) Finally, there was the radical anti-Humean like Dancy who argued that desires are not what ultimately motivate actions. Combinations of cognitive states are sufficient for action. Though, according to this theory, desires can be *ascribed* whenever the gap between the two representations motivates, this is not to claim that desires are what *ultimately* motivate.

What I want to suggest by lining up the different accounts like this is that all theorists agree on one thing, namely that the weakened version of the first Humean thesis is true:

- (3c) It is part of our concept of belief that beliefs can motivate only in the *presence* of a desire (widely construed)

The theorists disagree on whether the desire in question is the ultimate motivator or not, whether it has to be a natural desire or it could be caused by normative beliefs, whether desires can be identical with beliefs or not. But they all seem to find it counter-intuitive to suggest that one could be motivated to ϕ without desiring (in a wide sense) to ϕ – even a radical anti-Humean like Dancy. And, indeed, it is genuinely platitudinous that any agent will have a *goal* – will have something he *wants* to achieve, will *desire* (in the wide sense of the word) something – in so far as he acts.

How one might *further* characterise this desire is matter for substantial psychological theorising – going beyond the merely platitudinous.

Thus I propose that (3c) is what is genuinely platitudinous about the relation between belief and desire in the production of action. And a weak principle like this can, of course, be accommodated by the minimalist's notion of minimal belief. So, to conclude, the concept of minimal belief can accommodate all the psychological platitudes (when they are correctly formulated) and is therefore *no* mere decoy-concept.

5. *Answering the Australian challenge.*

Now, to recapitulate the dialectic of this chapter, it seemed at the beginning – due to the Australians – that the only way for the minimalist to block the standard argument for expressivism would be to engage in substantial philosophical debate about moral psychology. Thus, it seemed that if the minimalist wants to leave the expressivist outside the realism-debates for good, he will have to show why the standard argument for expressivism fails. And as we have seen he has at least three options to choose from if he wants to do this. The problem for the minimalist that then presents itself is that, no matter which option he chooses, it is clear that minimalism about truth and truth-aptness no longer gives us ethical cognitivism on the cheap. It takes *substantial philosophy* to defend either theory, the Australians argued.

But, in the light of the foregoing section – where a notion of minimal belief was defended – we must now ask whether there really is a problem for the minimalist here. First of all, one strategy one might adopt is simply to concede that if the minimalist wanted to argue that we could get rid of the expressivist once and for all merely on the basis of a purely platitude-based account of truth and truth-aptness, then he has been proved wrong. But, then also say, with Wright, that this was never the intention of Wright's minimalist. Thus Wright writes:

I have, I admit, presented no consideration which, strictly, imposes the minimalist way of looking at these matters. Rather, it seems to me that we are here concerned with issues which turn on theoretical advantage. And minimalism promises considerable advantages if it can be sustained'.⁷⁹

⁷⁹ Wright (1994c).

The theoretical advantages referred to here are, of course, the advantages of not having to give an account of logical inference between statements only some of which are genuinely truth-apt according to the expressivist – that is, the famous Frege-Geach-Hale problem.

But, what I want to argue in the light of the foregoing is that there now is *a much more direct way* of blocking the Australians' argument – now, that is, that we have defended a notion of minimal belief. More precisely, we may now ask *what notion of belief is the one featuring in the second premise (and conclusion) of the standard argument from expressivism and in the platitude (P)*. And what we will find is that we cannot make the platitude and the second premise true simultaneously while interpreting 'belief' consistently. Consider that the following principle is quite true and platitudinous:

(P*) If *s* is minimally truth-apt, then *s* can be used to express the content of a minimal belief

But, notice also that the following version of the standard argument for expressivism does not go through because the second premise is simply not true:

- (1) The states expressed by the sincere utterance of moral sentences are intrinsically motivational (when about the obligations of the subject himself).
 - (2*) No *minimal belief* is intrinsically motivational.
-
- ∴ The states expressed by the sincere utterance of moral sentences are not minimal beliefs.

That is, it is simply not true that no *minimal* belief is intrinsically motivational, since some minimal beliefs are moral beliefs and they are arguably intrinsically motivational (in the sense given either in the weak or strong internalist claims). So, if we consistently interpret 'belief' in the Australians' challenge as *minimal* belief, their argument will not go through.

On the other hand, if we try to make the second premise true – by using here the notion of a robust belief and claiming truly that no *robust* belief is intrinsically motivational – we have a problem with making a version of (P) true. That is, the

following is clearly not a platitude, since *s* might be truth-apt while only apt for expressing merely minimal beliefs:⁸⁰

(P**) If *s* is minimally truth-apt, then *s* can be used to express the content of a robust belief

A discourse's being minimally truth-apt is simply not sufficient for its being apt for expressing robust beliefs. But then there is a direct answer to the Australian challenge that does not rest on a willingness by the minimalist to engage in discussion of substantial issues in the philosophy of moral psychology. The notion of a minimal belief – that we have seen can be defended as a genuine notion of belief – allows us to block the Australians' argument without substantial philosophical commitments, since it deprives the Australians of the opportunity to make trouble for the minimalist through exploiting the platitude (P) and the standard argument for expressivism.⁸¹

As a final comment, let me suggest that what all this shows is that the minimalist – with his distinction between merely minimal belief and robust belief – can say exactly the same things as the expressivist can – with *his* distinction between *commitments* and beliefs.⁸² At bottom this may simply be a terminological quarrel⁸³ and what remains is, then, that the minimalist approach is immune to the Frege-Geach-Hale objection that haunts the expressivist. Therefore, it would seem that the minimalist has a clear dialectical advantage over the expressivist.⁸⁴

⁸⁰ This way with the Australian challenge seems also to be advocated in Wright (1998a) and (1998b).

⁸¹ There is a further, *ad hominem*, argument against (the Australian) Smith's suggestion that expressivism cannot be excluded from the debate by a new neutral, platitude-based, framework. That is, it seems in the light of the foregoing that Smith himself must believe that expressivism can be excluded on the basis of what is a mere platitude. Smith argues that weak internalism is platitudinous. That is, it is analytical that if *x* believes that he ought to ϕ , then – other things being equal – he desires to ϕ . Strong internalism, on the other hand, flies in the face of various phenomena of practical irrationality according to Smith. But, an expressivist must accept something similar to strong internalism. Since on his account the states expressed by moral statements *are* themselves motivational, it is impossible to have a moral attitude while not being motivated. So, we can reject expressivism – according to Smith – merely on the grounds that it fails to accept that weak and not strong internalism is platitudinous. Hence Smith ought not to have any problem with rejecting expressivism on the basis of platitudes.

⁸² For more on this distinction between beliefs and commitments, see for instance Blackburn (1984), chapter 6.

⁸³ In support of this claim, see Scanlon (1999), p. 64. "I have argued that if there are substantive standards relative to which judgements of a certain kind can be called correct or incorrect, then there may be no reason not to take them to express beliefs. But, if there are such standards, then the choice between these two interpretations may no longer make much difference: how much should we care about the difference between saying that these judgements express beliefs and saying that they express other attitudes for which there are clear standards of correctness?"

⁸⁴ See also Thomas (1997) for this point.

Chapter Three

Moral Minimalism and Moral Disagreement

Many philosophers have held that there are radically un-settleable disagreements over moral issues and have taken the phenomenon to have drastic consequences for our understanding of moral discourse. In particular, it has been thought to have nihilistic or (strongly) relativist consequences for our thinking about morality.⁸⁵ The putative examples of radical disagreements that are typically given are disputes over the moral status of strangers, foetuses and animals, euthanasia, the relative value of freedom versus equality and a host of other disputes over the relative weight of various values. Often these disagreements derive from differences between cultures, but of course they may be differences between sub-cultures or generations within a single culture or even just individuals with radically different moral sensibilities. Certainly examples seem many in today's pluralistic Western societies. In this chapter, I deal with how the minimalist can account for this phenomenon.

It may seem that there is a serious problem for the minimalist who wants to face the phenomenon of radical moral disagreement. The worry is one about accommodating three conflicting intuitions simultaneously. Firstly, we seem – *seem* – to have to accept that there are actual radical moral disagreements – disagreements that no further investigation of any kind *could* settle. Call this *the intuition about radical disagreement*. Secondly, we seem to have to accept that at least some of these are *genuine* disagreements – that is, disagreements of the form: A believes that *p* and B believes that *not-p*, where *p* and *not-p* are genuine contradictories. Call this *the univocality intuition*. Thirdly, as the minimalist argues, there cannot be moral truths that are in principle unknowable. Call this *the intuition about epistemic constraints on moral truth*.⁸⁶ The conflict becomes apparent by asking: How is a genuine disagreement possible where it is un-settleable, yet where it must be settleable?

The immediate temptation is, of course, to go relativist and to conclude simply that actual radical disagreements are merely *prima facie* disagreements. That is, give

⁸⁵ E.g. Mackie (1977) and Harman (1975, 1978, 1989) and Harman and Thompson (1996).

⁸⁶ The argument for the minimalist's acceptance of this intuition has, of course, been given in chapter 1.

up the univocality-intuition and say that the parties in the dispute are using the same *words*, e.g. 'right' and 'wrong', but with different *senses*. The suggestion might be that the senses of the words are in some way relativised – e.g. to a moral system, a person, a culture or a gender. However, as I shall argue in a later chapter, shedding the univocality-intuition amounts to preventing moral discourse from performing its essential social function.⁸⁷ The minimalist, I argue, must focus on the characterisation of radical disagreement – given that he is committed to epistemic constraints on moral truth (I take this for granted in this chapter) – in order to resolve the conflict between the three intuitions.

The structure of the chapter is as follows. I want, first of all, to pin down the phenomenon of radical moral disagreement by defining a notion of *an un-settleable moral dispute – or radical moral disagreement* (section 1). Secondly, I distinguish between *de facto* and *absolute* un-settleability (section 2). And thirdly, I shall through examining a set of examples attempt to provide a taxonomy of *types* of radical moral disagreement (section 3). Armed with these definitions and examples, I then turn to the question how the minimalist can reconcile the three intuitions.

I, first, explain how the minimalist can account for the univocality-intuition. How is it possible, on his account, for two people to genuinely disagree over a moral issue, yet have very different sets of basic moral principles (section 4). In this section, I also deal with a relevant objection – advanced by Folke Tersman – that radical moral disagreement cannot occur at all on purely conceptual grounds to do with the possibility of radically interpreting the disputants. The nub of my defence of minimalism is here a distinction between *core* moral platitudes and peripheral *moral* platitudes or principles.

Then I turn to characterising the intuition about radical moral disagreement. I shall reject two proposals for how to do this before settling on the minimalist's suggestion. The first, straight-forward, proposal is simply *that there are, or could be, absolutely un-settleable moral disputes* (section 5). Arguably, the moral error-theorist or nihilist bases his treatment of moral discourse on this claim. I shall call it Scylla, since the minimalist can only accept it on pain of inconsistency. But, I shall also argue

⁸⁷ See chapter 5 – the sections on the explanatory approach and relativism.

that the minimalist does not have to accept this claim, since our actual knowledge of moral disagreement cannot support it.

The second proposal for dealing with our intuition about moral disagreement is to shed it completely – i.e. say *that we have an a priori guarantee that all moral disputes are settleable* (section 6). This proposal – I call it Charybdis – amounts to the claim that moral discourse exhibits Cognitive Command. Underlying it is a strong form of optimism about the potential for improvement in our means of moral argument – a claim that something like effective decision-procedures can be provided for moral questions. I shall argue that – though it is not ruled out *a priori* that it can be achieved – this is too strong a claim to advance as an account of actual moral discourse.

Finally, I then settle on a third proposal that sails between Scylla and Charybdis. This proposal is that all we can claim about settleability in moral discourse in general is *that we, so far at least, have no a priori guarantee that all moral disputes will be settleable* (section 7). In other words, that we ought to be *agnostic* about the a priori status of claims about moral settleability. In effect this claim about moral un-settleability is equivalent to the combined claims that moral discourse is minimally truth-apt and that it does not exhibit Cognitive Command. Thus, I argue that minimalism can embrace this characterisation of the phenomenon of radical un-settleability and accept the other intuitions – about epistemic constraint and univocality – while remaining consistent.

1. What is an un-settleable dispute?

First of all, how are we to define 'un-settleable dispute'? Evidently, there is a trivial sense in which a dispute between two individuals may be impossible to settle that should not hold us up here where we are concerned only with un-settleability as it pertains to the objective status of morals. 'Un-settleable' can be taken to apply to a dispute when it is a case of two individual's never reaching agreement on some matter simply because of stubbornness on the side of one or both parties – an irrational unwillingness to face the agreed facts or to give in to what ought by his own lights to be a convincing argument. But, that tells us as little about the objectivity of morals as a similar kind of dispute within science tells us about the objectivity of science.

Perhaps many of the heated disputes about moral issues that actually take place today are like this – maybe they are to be likened to the disputes had within evolutionary science between the establishment of neo-Darwinians and religiously motivated 'dissenters'.

A more interesting, deeper, sense in which disputes can be said to be un-settleable is what will interest us in the following. Let me give a preliminary definition:

A dispute is un-settleable if and only if there is *no objective basis*
available to the disputants on which to settle it.

This is, of course, an empty formula as it stands. Two questions immediately arise. What is *an objective basis* on which to settle a moral dispute? And what does *available* mean here? To get a better grip on the notion of *an objective basis* (leaving out for now the requirement that it be available to the disputants), I now want in general terms to discuss the different ways in which an objective basis can exist on which to settle a moral issue.

Imagine that A and B are involved in a disagreement over a moral issue. A claims that p and B claims that $not-p$. The disputed p can here be a particular moral claim of the general form Ma . Or it might be a universal moral principle of the general form $(\forall x)(Fx \rightarrow Mx)$ where M is a moral predicate (and F is either a moral or a non-moral predicate and may be of high complexity) – what I shall call a *linking-principle* in the following.⁸⁸

A first, trivial, way of resolving a seeming dispute on an objective basis is by discovering that the disagreement is due to a simple failure of rationality on one or both sides of the dispute. If, that is, the disputed beliefs are justified by inference from evidence via inference-principles, there is the possibility that one or both disputants have made errors of reasoning – have been prejudiced in assessing evidence, have made a logical error and so on. Thus, both parties might actually agree on all the relevant evidence and principles of inference, yet come to divergent result. Let me in the following discussion for reasons of simplicity leave out any further mention of this way of finding an objective basis for the resolution of a moral dispute. (In effect, I shall assume that disputants are ideally rational in the following).

⁸⁸ I use Wright's term, see Wright (1992), p. 152.

But there are other ways in which there may be an objective basis to solve a moral disagreement. The question to ask now is whether (a) A's and B's beliefs are justified inferentially from other background moral and non-moral beliefs or whether (b) they are not so justified. Let me take the *inferential* case, (a), first.

(a) *Disputes over inferentially based beliefs*: Now, if A's and B's disputed beliefs are reached via inference from other beliefs, they cannot have reached their contradictory verdicts on the basis of the *same* evidence while correctly using the *same* principles of inference. Thus, if we disagree over whether Peter acted wrongly, we cannot agree both that Peter was lying and that lying is wrong (and that Peter's action had no further morally relevant features). Hence, there will have to be a further difference between A and B with respect to either (i) evidence or (ii) linking-principles.

(i) If we take the case of an *evidential dispute* first, then either the evidence is itself formulated in moral terms or it is formulated wholly in non-moral terms. If it is moral, then the process starts over and we must ask whether this moral background belief is inferentially justified or not. If, however, the evidence is non-moral, then there is the possibility that an objective basis for a resolution of the dispute may consist in the possibility of a resolution of these differences in non-moral background beliefs.

(ii) But, the initial dispute may also turn out to originate in an *inferential dispute* – a dispute over linking-principles. Now, again the disputed linking-principle may itself be inferentially justified in which case the process starts over again. (For example, the relatively specific moral linking-principle 'Lying is wrong' could be justified inferentially by a more general principle like 'Dishonesty is wrong' in conjunction with the belief that 'Lying is a form of dishonesty'. Thus, the search for an objective basis for a resolution of the initial dispute goes on as a search for an objective basis for resolving the background disputes about the more general linking-principles). But, of course, this process cannot go on forever – justification cannot proceed from an infinite regress of justifiers. At some point the chain of disputes must end up in a dispute over linking-principles that are not themselves inferentially justified. So, we turn to disputes over non-inferentially justified beliefs.

(b) *Disputes over non-inferentially justified beliefs*. We have now reached a type of dispute that we may call a *basic dispute*. That is, a dispute is basic if and only

if the disputed p is regarded by the disputants as not justified by further background beliefs of any kind (though, of course, there may be undefeated *entitlements* to hold the disputed p true). In particular, in a basic dispute the disputed p cannot be a verdict about a *derived* linking-principle – it must not be held to be justified by further background moral principles. And, of course, it cannot derive from a further evidential dispute over a non-moral issue. So, the question is now: How might an objective basis for the resolution of a moral disagreement exist given that the dispute is basic? There appears to be three ways in which a basic moral dispute could have an objective basis on which to settle it (remembering that we are here not concerned about whether this basis is available to moral subjects): 1) The world might settle it perhaps beyond our ken, or 2) moral epistemology might be intuitional, or 3) moral discourse might have necessary ties to other discourses such that moral disputes carry over automatically to these.⁸⁹

1) I shall say little here about the possibility that moral truth is evidentially unconstrained and, thus, that an objective basis for resolving all moral disputes might be 'out there' regardless of whether we can know this – i.e. that bivalence holds unrestrictedly. As I have already argued, the minimalist believes that moral truth is epistemically constrained.

2) But, even if moral truth is epistemically constrained, we may still have a guarantee that there is an objective basis for a resolution of basic moral disputes, if moral epistemology is intuitional. That is, we may be able to defend the claim that we have a faculty – much like we have a faculty for perception and, perhaps, mathematical intuition – that is somehow able, under conditions of normal functioning, to detect moral truths. If this is the case, then an objective basis for settling a basic dispute may consist in pointing to failures in the functioning of this faculty – that is, abnormality in either the *circumstances* or the *subjects* of moral perception – on one or both sides of the dispute.

3) Finally, it may be possible to establish that basic moral linking-principles do *not* float free of an anchoring in other discourses. In other words, that disputes about basic moral principles are necessarily connected to disputes within some other relevant discourse and, hence, that it may be possible to arbitrate basic moral disputes

⁸⁹ These considerations implicitly answer the first objection to the very idea of a radical disagreement

within this discourse – the discourse that morals disputationally supervene on. Thus, the claim could be that moral discourse 'disputationally supervenes' on some other discourse, to use Wright's expression.⁹⁰ For example, moral discourse might be thought to disputationally supervene on discourse about the will of God (Divine command theorists), or about the nature of human rationality (Kantians), or about empirical discourse about what makes people happy (naturalistic utilitarians). Hence, to take the first example, a radical dispute over some moral principle could become a dispute over how God wants us to act – a dispute which itself has an objective basis on which to be settled.⁹¹

So, to sum up these considerations, we have identified several possible ways in which there may exist an objective basis for settling a moral dispute: There could be flawed reasoning or false non-moral evidential beliefs on the part of one or both disputants. Or, the world might simply provide a fact, possibly beyond our ken, that makes the disputed belief true or false. Or, there could be an intuitional moral faculty and the malfunctioning of this faculty in one or both disputants. Or, moral discourse might disputationally supervene on some other discourse in which, in turn, all questions are guaranteed to be settleable.

Conversely, a moral dispute might turn out to have *no* objective basis for a resolution in two general ways. It may be down to a non-moral evidential dispute which does not itself have an objective basis for its resolution. Or, it may be that the dispute is basic (about an ultimate, affectively grounded, belief about a moral linking-principle), that moral truth is evidentially constrained and that it is neither the case that moral epistemology is intuitional nor that moral discourse supervenes on some other objective discourse.

Let me notice here in passing that, through these considerations, the contours of the minimalist's position has come within reach of our mental eyesight. The minimalist, crucially, remains agnostic about whether there is a guarantee of an objective basis on which to settle all basic moral disputes in either of the kinds just outlined while accepting that all moral truths are knowable. To see why the minimalist remains agnostic, recall that the basic minimalist idea is that the epistemology of

in Tersman (1998).

⁹⁰ See Wright (1992), p. 155.

morals is such that we are *entitled* to hold basic moral principles true – namely the moral principles that our affective nature make us spontaneously accept – without at the same time having incurred any further commitments to justify these (though, of course, our entitlements may be defeated). Hence, the minimalist claim is that morality is autonomous and *floats free* of ties to other discourses – there is no disputational supervenience on any other discourse. Morals is based on human *spontaneity* to use a Kantian expression. There is no moral *receptivity* and therefore no *sui generis* intuitional epistemology of morals. Rather, the claim is that moral discourse is ultimately grounded (though defeasibly) in our subjective natures – our spontaneous affective responses. Hence, the agnosticism about the guarantee of finding an objective basis to settle every moral dispute. As already mentioned, the question guiding the central parts of this chapter will be how to characterise actual radical moral disagreement in a way consistent with this minimalist account of moral discourse.

2. *De facto versus absolute un-settleability*

But, to return now to the question posed at the beginning of last section – what is a radically un-settleable dispute? – it must be noticed that the mere existence of an objective basis as described in the foregoing is not sufficient to make a dispute settleable. If, namely, the objective basis is itself beyond our ken, then its existence will not be able to help us to an actual resolution of the dispute. So, for a dispute to be settleable there has to be an objective basis in the sense explained above and this basis has to be *available* to the disputants.

This brings up the second question. What exactly does 'available' mean here? We can answer that some fact is available to an investigator if and only if the investigator possesses a method of gathering and processing evidence – or there is some feasible way of adding to or refining the investigator's methods – such that when using this method, the investigator *will* as a matter of fact (in a feasibly finite amount of time) come to know this fact. Hence, some fact is available if and only if it is

⁹¹ The implications of disputational supervenience for the possibility of defending moral minimalism are explored in chapter 6.

knowable. An example of an *unavailable* fact in this sense could be a fact about the exact temperature on this exact location on Earth exactly one million years ago.

With all these distinctions now in play, we have given some substance to the idea of an *un-settleable dispute* as a dispute where no objective basis for its settlement is available. We see now that the lack of availability may be down to one of two things. The fact that the objective basis – though existing – is radically beyond our ken or to the fact that it simply does not exist. But there is still a crucial ambiguity in the definition of an un-settleable dispute that I shall now remove by distinguishing between *de facto* and *absolutely* un-settleable disputes. The ambiguity lies in the word 'is' in the expression 'no objective basis is available'. This may be taken to mean that as a matter of fact – so far – no objective basis has been found to resolve a dispute. Or it may be taken to mean that no objective basis *could* ever be found on which to resolve the dispute. This distinction allows us to make both a weaker and a stronger claim about actual moral disagreement.

The weak claim is that, for some actual disputes within moral discourse, we know that we have so far not found any objective basis for a settlement in any of the ways pointed to in the previous section. Let me rephrase this by introducing a new term: Some actual moral disputes are *de facto un-settleable*. Let me here venture the opinion that there are actual moral disputes such that we know that we at present have no effective method of finding an objective basis for settling them. I suspect that this claim, that some disputes within ethics *as a matter of fact* end in a stalemate, is uncontroversial.

But, we can also make a stronger claim about actual moral un-settleability. Namely that there are disputes within ethics such that nothing *could* ever decide them. This, I say, is the claim that some moral disputes are *absolutely un-settleable*. That, for some actual moral disputes, we know that no objective basis for resolution *will ever be available*. This stronger claim – which crucially involves a claim about the impossibility of ever refining or adding to our methods of reasoning in a way that will make the dispute resolvable – certainly should be seen as more controversial. And there is clearly no immediate step from *de facto* un-settleability to absolute un-settleability. To anticipate, I shall later argue that the weak claim is true and compatible with the minimalist's position whereas the strong claim is incompatible

with the minimalist's position, but also false. But, first some examples of radical moral disagreement.

3. *Examples of radical moral disputes.*

So far, I have tried to arrive at an abstract conception of what a radical moral dispute can look like. I now want to move closer to actual examples of radical moral disagreement and divide the class of radical moral disputes into some further subcategories. In the foregoing, we have seen that (non-trivial) radical moral disputes derive either from a radical evidential dispute over a non-moral issue or from an inferential dispute over a basic moral linking-principle. I shall now distinguish three different types of dispute over moral linking-principles and I shall give some concrete examples of these together with examples of evidential, non-moral disputes that underlie moral disputes about particular cases.

To understand the distinctions I want to draw between different types of dispute over linking-principles, it is important to draw distinction within the class of moral judgements between *evaluative* and *deliberative or practical* judgements.⁹²

An *evaluative* judgement states that something is valuable or good or good-in-itself. We may here further distinguish between moral (impartial, other-regarding) and prudential (self-regarding) values – in the following the evaluations, I refer to, are to be understood as moral evaluations. Evaluative judgements come in thin versions, e.g. 'x is good' or 'x is valuable'. But also in thick or more specific versions with empirical content mixed in with the evaluative, e.g. 'x is chaste', 'x is just' or 'x is cruel'. X typically refers here to people, acts or institutions. Accepting a value is to accept a principle for evaluation of the form: If x is such and such then x is valuable/good/just/chaste *etc.* Part of adopting a value like this is to accept that certain natural features (of acts, persons, institutions) ought to have some bearing on how we act. Thus, one way to say what it is to have a value – say, the value of honesty – is to say that: One ought *other things being equal* to be honest.⁹³

⁹² I here follow Wiggins (1998), p. 95-96 and p. 155.

⁹³ Thus, I am here adopting the view that value-judgements can be seen as general judgements about *pro tanto* obligations. See Wiggins (1998), p. 96n6.

Notice now also that people may have different lists of ultimate values. Some may have a list of only one ultimate value as the utilitarians do – utility. But, most often, of course, people accept that there are a plurality of mutually irreducible values, including both prudential and moral ones. Thus, people can value pleasure, understanding, deep personal relationships, life, honesty, freedom, justice, autonomy, chastity without thinking that any one of these can be reduced to – subsumed under – one of the others.⁹⁴

A *deliberative or practical* judgement, on the other hand, is a judgement designed to guide action directly. It is a judgement of the form: 'S (morally) ought or must, *all things considered*, ϕ in circumstances C'. Typically, reaching a deliberative verdict about a particular course of action in a particular situation involves bringing *different* values into play. Thus, in a particular situation we might face a conflict between the value of benevolence (i.e. not harming others) and the value of honesty. Say that we have to tell the truth to someone whom it would hurt. Thus, we might accept that two evaluative principles ought to guide us in giving a deliberative verdict in such a situation: 'One ought *other things being equal* to avoid harming others' and 'One ought *other things being equal* to be honest'. But, these principles do not by themselves determine a verdict about how one *must* act in the particular situation – what one ought to do *all things considered*. What verdict is given will depend on how we believe the two values should be weighed against each other. So, it will depend on the acceptance of further, more specific, principles that tell us what to do in situations where these values conflict.⁹⁵

This observation leads me to make an important point before turning to more concrete examples. We may expect that the scope for disagreement about these more specific principles – principles about exactly when things are not equal with respect to the various *pro tanto* obligations, about when to make exceptions from the general rule – is much larger than for the general valuational principles. Thus, we may conjecture that disagreements about general evaluational principles or particular evaluative judgements are not as likely to be radical as disagreements about

⁹⁴ For a plausible list, see Griffin (1996), p. 29-30.

⁹⁵ What principles we, then, arrive at will depend on what our *attitude* towards the values is – in particular, whether we believe that we must *respect* or *promote* values. I shall say more about this distinction in the following chapter on moral dilemmas.

deliberative or practical principles or judgements – verdicts about the overall significance of values in specific situations. This seems also to be Wiggins' view.

‘So I end by saying that, where matters turn on the application of strictly valuational predicates in a fully determinate context, the question of truth will both arise and stay around for an answer; but where practical judgement comes into consideration – and with it the question of *what* categories of evaluation and modes of attention we are to bring to bear upon a given practical question – we must expect truth to be much more problematical. Which is not to say that it is out of the question for truth to be attained.’⁹⁶

With this conjecture, let me now suggest that there are three general ways in which a basic dispute about a linking-principle may take shape. We may have a dispute about the exact content of a single evaluative principle. We may have a dispute about what belongs to our list of values – what basic 'categories of evaluation' to adopt in the first place. And, finally, we may have a dispute over basic principles for weighing values against each other in particular situations where we need to make a practical verdict. I shall now give examples of these three types of moral disagreement together with examples of apparently un-settleable disagreements deriving from un-settled questions about non-moral information.

1) A basic moral disagreement may turn out to hinge on a disagreement about a basic linking-principle connected with a single moral value-concept and, hence, end up as a disagreement about the extension of the value concept. Thus, we can envisage a disagreement about what it takes to be courageous. A might accept the principle: 'For all x , if x faces danger to degree n or higher in order to achieve something worthwhile, then x is courageous', while B might subscribe to the principle: 'For all x , if x faces danger to degree $n+1$ or higher in order to achieve something worthwhile, then x is courageous'. Hence, their particular verdicts about courageousness may contradict each other – indeed B might think that A calls certain non-courageous, or even cowardly, acts courageous.

Another more interesting example here is the concept of a moral subject. People disagree, for example, about whose well-being we are to take into account or who we are to award rights – where to draw the boundaries of the extension of ' x is a moral subject'. More specifically, we may imagine somebody defending the view that

only the well-being of human beings (the biological species) are a matter for our moral concern coming up against somebody who holds that the well-being of all animals capable of pain (say defined by their having a central nervous system) should be our concern. A similar dispute could take place with regard to the moral status of fetuses.

To look at the example of a dispute over whether to include animals in the extension in more detail, we may see this as a dispute over which moral linking-principles to accept. One disputant accepts the principle: For all x , if x is a human being, then x is a moral subject. The other disputant may, then, accept principles like: For all x , if x is a sentient being (as defined in biological terms), x is a moral subject. Or, more radically: For all x , if x is a living being, then x is a moral subject. Again we seem to have identified a type of disagreement about a primitive principle that may be basic.

2) A radical moral dispute can also come about due to one party accepting moral values as basic that the other party simply rejects (or treats as less than basic). A famous example is here Oscar Wilde's proclamation during a trial when asked whether he did not find a passage of his own writings *obscene*. He answered: "Obscene" is not a word of mine.⁹⁷ By rejecting the very use of the thick evaluative concept of obscenity, Wilde did not directly disagree with the judge over whether something was obscene or not. With thick ethical concepts, if one does not agree with the normative principle that it encapsulates and was used to formulate, one has to reject the term completely (unlike with thin ethical terms). But, by not accepting the disvalue of obscenity, Wilde reached a (negative) verdict on whether, all things considered, it is morally wrong to publish the kind of writing that he had published. And this verdict *did* contradict the judge's verdict on this question. So, we see how including different values in one's list of ultimate values can lead to radical disagreements.

We may expect that clashes between radically different cultures or different eras within one culture will provide many examples of clashes between different moral categorisations – systems of thick ethical concepts – like this. An, often picked on, example of this is the dispute over what types of sexual behaviour are permissible.

⁹⁶ Wiggins (1998), p. 351, see also p. 155.

There is a Western Christian tradition according to which there are *special* moral principles guiding our sex-lives. Thus, *chastity* is seen as a basic value and sexual acts of a certain form are consequently deemed morally wrong – roughly those sexual acts that are not aimed at procreation and family-making like masturbation, gay sex, sex with contraceptives and sex outside marriage. Against this there is the – surely more widespread – view that there are no (true) moral principles specifically about sexual conduct. According to this view only general moral principles for guiding action – about concern for the well-being of others, about respect for others and about honesty, for example – ought to guide our sexual behaviour. So, again, a follower of the tradition mentioned (who may of course accept all the general moral norms too) may end up in a radical dispute with a modern free-thinker about sexuality where everything hinges on one party accepting a value-principle that the other simply does not accept.

In general, there will be possible disagreements of this second type whenever disputants accept different sets of basic values – different lists of goods-in-themselves. A special case here is the utilitarian who accepts a very short list of values consisting of only one: utility, happiness or something similar. Thus, the utilitarian will potentially experience radical disagreements with anybody who accepts a more pluralistic account of values. A number of examples of radical disagreement of the second sort can thus be generated by giving attention to the many ways in which utilitarians have come up with what value-pluralists take to be 'counter-intuitive' answers to specific moral questions.

3) The third type of radical dispute over linking-principles arises in cases where there is actually agreement between the disputants about both the relevant non-moral beliefs and the list of basic values, but where there are differences over the exact contents of the linking-principles that determine how important the values are relative to each other in a particular situation. The third example of a basic dispute, thus, derives from the fact that this difference in linking-principles will cause the disputants to apply the value-principles differently when faced with particular circumstances in which to make a practical verdict – that they will weigh moral values

⁹⁷ As reported by B. Williams (1996), p. 29.

differently against each other in particular cases where there are more than one value in play.

Think here about the well-known moral debate about euthanasia. We may imagine that the parties to a dispute about a particular case of euthanasia agree about there being two moral values at stake. They both accept, on the one hand, a norm stating that one ought, other things being equal, to relieve the pain of others (based on the value of benevolence) and, on the other hand, a norm stating that one ought, other things being equal, not to deliberately take (innocent) lives (based on the value of life). But imagine also that the disputants disagree about which value to give priority to in the particular situation.

One popular way to describe the possibility of basic disagreement of this kind is to say that it arises from the *incommensurability* of the two concerns involved. This claim may come in two strengths – as a claim about full or merely partial incommensurability. The claim of full incommensurability seems to be indefensible, though. When two value-norms conflict in a range of cases, there are typically sub-groups of cases where one concern clearly outweighs the other. Thus, thinking of the example of euthanasia, consider a case where someone is in severe pain due to serious illness, but where the prospects of speedy recovery are reasonably good. Here a concern for removal of pain – compassion – seems clearly outweighed by the value of life. This contrasts with the case of a terminally ill patient who is in severe and unrelievable pain to an extent which renders the patient's life devoid of all the qualities that we value as part of a good life and who expresses his or hers desire to die. Here – at least in my mind – compassionate killing is *clearly* morally permissible.

Thus, the interesting phenomenon is that of *partial* incommensurability. The idea is here that – flanked by cases like the ones just mentioned where one or the other norm clearly takes priority – there is a class of cases where neither concern clearly dominates. Or to put this more carefully, the boundary between the cases where one principle holds sway and cases where the other holds sway is vague. Crucially, it is about these borderline cases that disputants will tend to have different linking-principles.

To give an example of the linking-principles that could be involved in a radical dispute over the euthanasia-case, consider the following principle accepted by one party to the dispute: For all x, if x has severe and untreatable pains due to (probably)

terminal illness with a chance of recovery of less than 5 %, and x has asked to have an end to his life (and so on...), then a concern for relieving the pain of x outweighs a concern to protect x's life. Consider also the opponent who thinks that 5 % chance of recovery is enough to keep up hope and consequently accepts the principle that: For all x, if x has an unbearable and untreatable pain due to terminal illness with a chance of recovery of less than 1 %, and x has asked to have an end to his life (and so on...), then (and only then) a concern for relieving the pain of x outweighs a concern to protect x's life.

Such differences at the level of basic linking-principles (for weighing values) can, of course, result in the disputants giving contradictory practical verdicts about a limited range of cases. As mentioned above, many actual radical moral disagreements appear to be boundary-disputes between different concerns – values – of this kind. Whereas people often – at least *within* communities – share a list of values, a moral vocabulary, and agree about all the non-moral facts of the situation, they end up disagreeing about how to accord weight to these values in particular situations.

4) Finally, a radical moral dispute may originate in a disagreement about the non-moral evidence where there is no way of settling this evidential disagreement. Trivial examples of this kind of un-settleability can be constructed by thinking about moral judgements about persons or acts in the distant past (or less plausibly distant regions of the Universe) – where there is no longer, or there never was, a chance to get evidence about the exact non-moral circumstances.

But, for a more relevant example, think of two disputants disagreeing over the permissibility of euthanasia while agreeing that the main moral concern is to maximise goods like well-being and respect for life. Imagine that one disputant makes the empirical prediction that allowing the deliberate taking of lives will place society on a slippery slope towards a society in which life is *generally* not respected – hence a society riddled with mistrust and insecurity where the well-being of all citizens is diminished – and on this basis rejects allowing euthanasia. The other disputant denies this empirical claim and, goes on to accept euthanasia. We may here imagine that the underlying empirical dispute is of such a great complexity that it is in principle impossible for humans to resolve it. (Perhaps bringing about the data that could give a resolution would itself involve so much disruption in society that it would no longer be the same kind of society – hence no longer a proper test-case). Hence, there may

be no objective basis available for the participants on which to settle their moral dispute due to the impossibility of retrieving the relevant non-moral evidence.

Or, as a second example, imagine two disputants disagreeing about the permissibility of treating members of a certain species of animal in certain ways – say raising them under poor conditions for human consumption. They may agree that we should seek to prevent pain wherever it occurs – say, they may both accept a hedonistic utilitarian concern for maximising pleasure – yet disagree about whether a certain species of animals are capable of experiencing pain (described in some precise biological terms), hence ending up with opposing verdicts. Again we have a radical moral dispute caused by what is ultimately a lack of ways to resolve a non-moral – empirical – disagreement.

However, it seems clear to me that disputes generated in this way are not essentially *moral* disputes – though, as a matter of fact, many questions that are *called* 'ethical' in public debate today are of this kind. The radical disputes of this kind are not moral – I argue – because the disputants agree about all moral principles involved. They only end up making different particular moral verdicts because they disagree about the empirical data to which they apply the principles. Thus, of the four ways in which moral debates may turn out to be irresolvable, it is only the three first – types of basic disagreement about moral linking-principles – that can properly be called *moral* disputes.

Before going on, I want to add a comment about the potentially misleading simplification of matters that I have been guilty of in the foregoing. Generally, the examples, I have given, have referred to very simple and very general universal linking-principles. I now want to make the point that linking-principles – especially of the third kind – can be of huge, perhaps indefinite, complexity and specificity. Indeed, as Hare comments, they may not even be formulable in words.⁹⁸ We cannot make explicit our moral sensibilities once and for all in a set of simple and general principles – like 'Lying is wrong'. In practice, we have to specify in much more detail the antecedent of such inference principles to make them principles that we can sincerely adhere to. We do not carry around a simple and easily formulable moral sensibility.

⁹⁸ Hare (1963), p. 46-47.

Rather, we make explicit as much detail of this sensibility as is needed in particular contexts – i.e. as we engage in different moral deliberations and discussions.

To take an example of how real linking-principles take on a much more complex character in the context of actual moral problem, consider again the linking-principle I advanced as an example of a weighing-principle in the euthanasia-case. For all *x*, if *x* has severe and untreatable pains due to (probably) terminal illness with a chance of recovery of less than 5 %, and *x* has asked to have an end to his life (and so on....), then a concern for relieving the pain of *x* outweighs a concern to protect *x*'s life (and allowing *x*'s euthanasia is right). The revealing part of this principle is, of course, the insertion of "and so on..". In order for us to sincerely adopt this principle, the antecedent would have to be specified in much larger detail. Many relevant considerations can be added to the list incorporated in the antecedent. There would have to be conditions ruling out certain possible wider consequences of allowing the setting aside of the norm prohibiting intentional killing of the innocent. There would have to be no severely bad consequences to society as a whole. There would have to be no threat of the slippery slope mentioned earlier. There would have to be no severe bad consequences for the doctors carrying out the killing due to the fact that they have to dismantle what is (hopefully) a deeply seated disposition not to kill – no severe bad consequences for the doctors' self-conception and the general conception of their role in society. There also have to be further conditions about the patient to make the principle plausible. Conditions about the exact degree of pain he is in, and the degree to which he is deprived of things that make his life worth living. Conditions about the patient being in a clear and lucid state of mind when giving his consent. Perhaps even conditions about the conception of euthanasia among the closest relatives should be included in the antecedent.

All these are relevant considerations and need to be taken into account when formulating one's universal linking-principles about euthanasia. And, of course, we can never be sure that the list is complete – we never know that we have taken everything relevant into account. As we consider new hypothetical cases, new provisos have to be added. Moral deliberation is in this sense *open-ended*. The principles we actually formulate in order to defend the claims we make in moral

debate are thus always provisional. The point is that we can make them as specific as we need to in the context of any particular dispute.⁹⁹

4. *The univocality of radical moral disagreement and the core platitudes.*

As noticed in the introduction to this chapter, one way to respond to the phenomenon of radical disagreement is to claim that there is only the *appearance* of radical moral disagreement – that all seemingly radical disputes are cases where disputants are using the same words with different meanings and, hence, speaking past each other rather than genuinely disagreeing. Against this proposal, I argue that we have a strong intuition about the univocality of radical moral disagreements. It seems readily intelligible for me to genuinely disagree over a moral issue with someone despite the fact that I accept radically different *mores* – set of basic moral beliefs about principles – from the *mores* of my opponent. I now want to show how the minimalist can explain this intuition by giving an account of what it takes to master moral concepts. And I want to use the explanation to answer an objection to Wright's minimalism posed by Folke Tersman based on arguing against the univocality intuition.

Thus, the claim I want to defend is that it is a feature of moral discourse that basic disagreement – disagreement about basic linking-principles – does not threaten univocality. Notice first, that this is a feature that moral discourse does not share with discourse about basic arithmetic. If I were to disagree with someone over a basic arithmetical truth – say the truth of a sentence like '7+5=12' – then communication would break down immediately. The only way to make sense of the other disputant would be to re-interpret his words as having a sense different from the sense my words have. With moral disagreement, on the other hand, we appear to be able to

⁹⁹ The point, I am making here, I take to be related to the point made by McDowell (e.g. McDowell (1998), p. 200-203) that it is not possible to 'disentangle' the cognitive or receptive and the evaluative element in our competence with an evaluative concept. That is, we may not be able to spell out exhaustively in non-moral terms under what conditions a given value-term applies. This point in McDowell can be recognised without accepting his further claim that one cannot truly acquire the evaluative belief without the corresponding motivation. For a recent discussion of this, see Blackburn (1998), p. 97-104. Blackburn introduces the useful term 'shapeless underlying class' (about the class – specified in non-moral terms – membership of which licenses the ascription of an evaluative predicate) to signal that we may not have a simple, formulable, non-moral description of the class of cases to which a given evaluative term applies.

retain a genuine disagreement – a focus for further discussion – even when the disagreement is deep, even when basic linking-principles are at stake.

To explain this feature of moral discourse, the minimalist can now claim that being competent with moral terms does not involve adopting any *particular* set of moral linking-principles. All that is needed to use moral terms correctly is compliance with a set of *core platitudes* – analytical truths – surrounding moral concepts. Thus, people that comply with these conceptual truths in their use of moral terms will be able to disagree intelligibly across even radically different *mores*. Contrast here what we might call *peripheral platitudes* – generally accepted moral standards like 'Lying is wrong', 'Helping others is doing good' or 'One ought to be chaste'. These are platitudes in the sense that they are (or have been) treated as common-places in particular cultural contexts. But, their acceptance is not necessary for mastery of moral terms.

Let me say some more about what these core platitudes are that surround moral concepts. They are, of course, all those judgements that someone who masters moral concepts cannot – on reflection – deny while remaining intelligible. Michael Smith has provided a list that captures many of the basic features of competence with moral concepts.¹⁰⁰ I have been inspired by this in the following.

First of all, it is a conceptual truth that moral judgements are *practical*. That is, it does not make sense for someone to say, for instance, 'To ϕ in these circumstances is the right thing to do, but I do not see the least bit of (even pro tanto) *reason* to ϕ now – and I am not in the least *motivated* to ϕ now either'. Thus, it is platitudinous that moral judgements are judgements about reasons and that they are accompanied with a certain degree of motivation. (It should be added here that the practicality platitudes only make explicit our competence with moral terms that are 'inherently action-guiding' to use Hare's terms – what is more often called thin ethical judgements.¹⁰¹ Thick moral judgements – like 'This is the chaste thing to do' or 'This is blasphemy' – are only 'contingently action-guiding'. That is, only action-guiding for those who accept the norms that they encapsulate. Others may intelligibly remain

¹⁰⁰ Michael Smith (1994), p. 39-41. Hare (see e.g. his (1999), essays 1,2 and 6) makes a similar point. He speaks about the 'logical properties' of moral judgements – their universalisability and prescriptivity – that can remain invariant across cultures and thus help us avoid moral relativism.

¹⁰¹ Hare (1999), p. 23-24.

unmoved while judging that these terms apply in particular situations – though they may prefer to stop using the word altogether).

A second cluster of core platitudes concern *moral debate* – the procedures for dealing with moral disagreements. First of all, it is a conceptual truth that contradictory moral statements really are contradictory – express contradictory moral views. That is, moral judgements have a pretension to be objective and, hence, moral differences issue in genuine disagreements. Further, it is understood by any competent user of moral terms that moral questions are debated in a rational manner. Moral particular claims ought to be justified by moral principles. And it is understood that we can attempt to resolve moral disagreements by engaging in a process of searching for a reflective equilibrium between our moral principles and our considered particular moral judgements about real and hypothetical cases.

Thirdly, it is platitudinous that moral properties *supervene* on natural properties in the sense that, for instance, two acts that are alike in all relevant natural respects are also alike in all moral respects. (I shall say much more about this platitude in chapter 6, so I shall leave it for now).

Finally, there are platitudes that serve to set moral judgements apart from other kinds of normative judgement. Thus, it is platitudinous to say that moral judges must take all persons involved in a situation into account – and accord them *equal* concern and respect – when giving moral verdicts. Moral judgements also tend to seek to improve the lives of human beings – to make them flourish.

This list is not exhaustive, but it can serve to make the following point. In so far as people with different *mores* share moral concepts in the sense that they share the type of practice with the concepts that are dictated by the above core platitudes, they will be able to disagree despite radical differences in basic moral outlook. In particular, they will be able to identify each other's moral judgements as moral due to their practical consequences. And they will be able to debate with each and seek a consensus in moral beliefs due to the platitudes about moral debate – and about the objectivity of morals and the call for principled justification.

This feature of moral discourse that it can encompass radically different moral views within one – that sets it apart from discourse about arithmetic, as we saw – is

widely recognised in meta-ethical debate.¹⁰² I now want to exemplify this account of moral terms as having a kind of ecumenical feature – a reaching across different starting-points – built into it by employing it to answer an objection to the very idea of (mere) minimalism advanced recently by Folke Tersman.¹⁰³ He suggests that methodological restrictions on radical interpretation ensures exactly that any apparent radical dispute (a cognitively blameless dispute) will turn out to be no real disagreement at all – that disputants involved in apparent radical disputes are merely speaking past each other.

The argument in a short version goes as follows: If two disputants hold the same sentence (under the same interpretation) true and false respectively, and if neither has a cognitive shortcoming, then (since the sentence cannot be both true and false under a single interpretation) one of the disputants must have made an *inexplicable* error with regard the disputed sentence. What could have explained the error would have been some false evidential belief, but this is ruled out by the claim that neither disputant is to blame cognitively. But, the argument goes on, a plausible principle for interpretation – a version of the principle of charity – states that we should always avoid ascribing inexplicable errors.¹⁰⁴ When we apparently have to ascribe an inexplicable false belief, we should instead change our theory of interpretation such that we can ascribe different, *true*, beliefs instead. Thus, Tersman concludes that a radical dispute over some sentence will really be a case of there being no disagreement at all – the participants, that is, will simply not be negating and affirming the *same* proposition when they use the sentence.

So, according to Tersman, the very idea of absolutely un-settleable disputes is incoherent because such disputes are impossible to interpret radically. Hence, failure of Cognitive Command is ruled out on conceptual grounds. There is more than one

¹⁰² At least among anti-realists. Hare appears to be the first to notice it, but many have formulated versions of it, Hare (1952), p. 146-9. Smith (1994), p. 32-35 recycles Hare's argument as part of an argument against the Cornell – ontological naturalist – realists. Horgan and Timmons' (1992) give a Twin-Earth argument against the ontological naturalist that makes a very similar point in a different terminology. (The similarities between Hare and the later arguments by Horgan and Timmons are noticed by Hare in Hare (1999), p. 82-86.) The possibility of *cross-mores* genuine disagreement – the possibility of radical interpretation of the deliberative judgements of foreigners with a radically different moral outlook – is also defended in Wiggins (1998), section 9, p. 116-18. And by Price (1988), p. 176-77.

¹⁰³ Tersman (1998).

¹⁰⁴ This approach to radical interpretation is, of course, that of Donald Davidson. See e.g. Davidson (1984), p. 125-139.

way to defuse this argument against the coherence of minimalism. Thus, it appears to presuppose that the principle of bivalence holds true a priori in morals – a point not agreed by the minimalist. Also, the argument appears to presuppose that the minimalist must characterise actual moral disagreement in the strong way envisaged earlier – as cases of absolute un-settleability. I shall argue in the next sections that the minimalist need not accept this.

But, there does seem to be a general worry that has to be answered on behalf of the minimalist. The problem is that we may – at least in the abstract – be able to imagine a dispute between people that are immersed in life-forms so different that it becomes difficult to claim that they share enough beliefs to be able to agree in meanings. Essentially, Tersman is arguing that Wittgenstein's famous observation also applies to moral discourse: 'If language is to be a means of communication there must be agreement not only in definitions but also (queer as this may sound) in judgements'.¹⁰⁵ In the particular case of moral discourse, we can imagine two cultures with very different sets of ultimate moral principles – peripheral platitudes – due to differences in the (educated) affective natures of the two peoples. In this type of scenario, we might want to say that we cannot give common meanings to deliberative terms like 'ought to or must (all things considered)', 'may', 'right' or 'morally permissible' or valuational terms like 'good', 'honourable', 'chaste'.

But, it is clear that the above considerations about the ecumenical character of moral concepts contain an answer to this objection. They amount to the claim that Wittgenstein's dictum does not apply in the moral context. Notice that Tersman's conclusion is absolutely general – that there are no radical disputes whatsoever. What I want to argue is that, whereas we may have to agree that *some* cross-cultural differences may be impossible to interpret as proper disagreements, we should certainly not agree in all cases. In particular, we can assign cross-cultural meanings to terms that connect more or less directly with action. As noticed above, *deliberative* – or more generally *thin* – moral terms are inherently action-guiding. This much is guaranteed by the practicality-platitude. And this very feature of moral terms can allow us to interpret across different basic moral outlooks.

¹⁰⁵ Wittgenstein (1951), §242. For doubts about whether Wittgenstein intended this dictum to apply to the moral language-game too, see Hare (1999), p. 61.

To see how this might be possible in more detail, consider that deliberative verdicts like: 'All things considered, x ought to ϕ ' will tend to issue in a characteristic range of behaviour in the maker of the verdict: negative reactions to x if he fails to ϕ , such as anger, verbal criticism, shunning, forms of punishment, and positive reactions of praise and encouragement if x ϕ -es. And there will be feelings of shame or guilt in x and perhaps attempts by him to apologise and make amends, if he himself makes the verdict, yet fails to ϕ . These fixed behavioural patterns can form the basis for a cross-cultural interpretation. It does not matter that in the different cultural contexts the verdicts are made with regard to different actions or when x is in different situations, nor that the verdicts are justified by radically different sets of – affectively grounded – background moral principles. The action-guiding and motivating aspect of these verdicts – ensured by compliance with the core platitudes – will ensure the possibility of interpreting people from radically different moral cultures as genuinely in disagreement over what to do.¹⁰⁶

In sum, Tersman's conclusion about the impossibility of univocal radical disagreement is simply too strong, ultimately due to its failure to take into account the special features of moral discourse – the distinction between core and peripheral platitudes. The minimalist *can* hang on to the univocality intuition. But the question then becomes: How should he make explicit the intuition about radical disagreement in a way that avoids the inconsistency (given the intuition about epistemic constraints on moral truth) referred to at the outset of this chapter?

5. First proposal for characterising actual moral disagreement – Scylla

As a first proposal for how we might characterise actual moral disagreement, consider the straight-forward suggestion that:

P1 Some moral disputes are (or could be) absolutely un-settleable.

This claim is inconsistent with moral minimalism – hence, he must avoid it like Scylla. However, it is also indefensible, as I shall argue. First, to see why it is inconsistent with minimalism, let me make some observations about the notion of an absolutely

¹⁰⁶ A similar argument is offered in Wiggins (1998), section 9, p. 116-18, and Price (1988), p. 176-

un-settleable disagreement – and the related notion of a cognitively blameless disagreement – and epistemic constraints on truth.

The claim that a dispute is absolutely un-settleable is an *epistemological* claim about the limits of our idealised powers to decide the matter in dispute. It is clear that this epistemological claim must not be equated with a further metaphysical claim, namely the claim that *the matter itself* is un-settleable. Or in other words, that there is *no and never will be any* objective basis on which to settle it – that the world itself is indeterminate with respect to the disputed issue. The latter claim is obviously a stronger, metaphysical, claim.

But, notice also that if the notion of truth applicable to sentences of the discourse is epistemically constrained, that is, if something like the following principle is true:

EC: For all p belonging to the discourse, p is true if and only if p is knowable

then epistemological and metaphysical indeterminacy are no longer distinct phenomena. That is, epistemological indeterminacy – for some p , $\text{not}(p \text{ is knowable})$ & $\text{not}(\text{not-}p \text{ is knowable})$, i.e. definite failure of completeness – becomes equivalent to metaphysical indeterminacy – $\text{not-}p \text{ is true}$ & $\text{not}(\text{not-}p \text{ is true})$. If all facts in the domain are knowable, then evidence that neither p nor its negation can be known is evidence of a truth-value-gap.

(1) $\neg \diamond Kp \ \& \ \neg \diamond K\neg p$	Failure of completeness
(2) $Tp \leftrightarrow \diamond Kp$	EC

$\therefore \neg Tp \ \& \ \neg T\neg p$	Truth-value gap

But, since it is inconsistent to accept truth-value gaps like this – the conclusion collapses into the contradiction $\text{not-}p$ and $\text{not-not-}p$ (at least under the assumption of the validity of the standard disquotational scheme: $Tp \leftrightarrow p$) – this is on the face of it bad news for anyone favouring the first proposal for characterising actual radical moral disagreement while accepting EC.

Let me in passing relate the notion of an un-settleable dispute to the notion of a cognitively blameless dispute. Under the assumption that EC holds for moral truth,

77. In chapter 6, I shall use similar considerations against the ontological naturalist.

the first proposal could be reformulated using the concept of a cognitive blameless dispute as:

P2 Some moral disputes are (or could be) such that neither participant has a cognitive shortcoming.

Of course, this reformulation is equally inconsistent. But, notice also that in a discourse where truth is *not* epistemically constrained, there is no direct move from the claim that a dispute is absolutely un-settleable in the way defined above to the claim that neither disputant has a cognitive shortcoming.¹⁰⁷ The possibility of evidence-transcendence of truth takes away the reasons to doubt that bivalence holds within the discourse.¹⁰⁸ Hence, the disputed claim will be guaranteed to be determinately either true or false and, therefore, at least one of the two disputants will trivially (albeit unknowably) have a cognitive shortcoming, namely in so far as he has the very belief that *p* or the belief that *not-p*.¹⁰⁹ If, on the other hand, truth is epistemically constrained in a certain context the fact that a dispute is absolutely un-settleable *does* imply that it involves no cognitive shortcoming on either side of the dispute. In particular, we have no longer any guarantee that A's belief that *p* or B's belief that *not-p* constitutes a cognitive shortcoming, since we have no longer any guarantee that bivalence holds true in the moral domain.

But, then we can also see that the point made here – that Scylla is to be avoided by minimalism on pain of inconsistency – is essentially the upshot of Shapiro and Taschek's¹¹⁰ argument considered in the first chapter. To remind ourselves: Shapiro/Taschek's argument was a *reductio*. They proved that the assumption that a particular dispute involved no cognitive shortcoming leads to inconsistency, given EC. Hence, they arrived at the following result (assuming EC):

¹⁰⁷ In passing, Tennant (1997, p. 51-53) seems to think that satisfaction of Cognitive Command is at odds with knowledge-transcendence. He, thus, says things like: 'satisfying Convergence [a principle which he takes to be in all relevant respects like Cognitive Command, see p. 52] makes it rather unlikely that the notion of truth could ever turn out to be a realist one' (p. 53). That is, the Cognitive Command test can only be a criterion that applies to discourses apt for an epistemically constrained notion of truth. As is implicit in the following considerations, I disagree with this interpretation of Wright and think – quite the opposite of Tennant – that knowledge-transcendence, since it motivates acceptance of bivalence, ensures satisfaction of Cognitive Command.

¹⁰⁸ I am ignoring here indeterminacy caused by semantic and other types of vagueness.

¹⁰⁹ At least given a thin interpretation of 'cognitive shortcoming' as this was explained in a previous chapter.

¹¹⁰ Shapiro/Taschek (1997).

It is a priori that, for any dispute, it is not the case that there are no cognitive shortcomings on either side of the dispute.

What the above considerations have shown is that this result might as well be reformulated using the concept of an absolutely un-settleable dispute. Given that, as I argued earlier, claiming that a dispute is absolutely un-settleable is the same as claiming that there may be no cognitive shortcoming involved in the dispute *when truth is epistemically constrained*, Shapiro/Taschek also give us this further result (again assuming EC):

It is a priori that, for any dispute, it is not the case that it is absolutely un-settleable.

To sum up: Given EC (accepted by the minimalist), both proposals, P1 and P2, for characterising the depth of actual moral disagreement collapses into inconsistency. With this result in, it is now time for a strategic reassessment. The question is how the minimalist can avoid the inconsistency. There would appear to be three possible responses:

1) The minimalist could avoid the inconsistency by conceding that moral truth is not epistemically constrained. But, of course, the minimalist does not accept this and cannot take this option without ceasing to be a minimalist.

2) The minimalist could follow Wiggins and give up on a genuine notion of truth in morals – and again cease to be a minimalist. Wiggins appears to do this, when faced with the inconsistency. He accepts that we *can* come up with cases of absolute un-settleability. (He seems to be thinking of cases of tragic dilemmas):

‘It seems that in the sphere of the practical we may know for certain that there exist absolutely undecidable questions – e.g. cases where the situation is so calamitous or the choices so insupportable that nothing could count as *the* morally reasonable answer.’¹¹¹

And he also appreciates the point that this leads to inconsistency (even for the intuitionist). But, he chooses to see this as a reason to think that the prospects of a strong or 'unrestricted' moral cognitivism are dim. He states:

‘If we insist upon the actuality of some absolute undecidability in the practical sphere, then we shall burst the bounds of ordinary, plain truth.

¹¹¹ Wiggins (1998), p. 130. See also, p.174-6.

To *negate* the law of excluded middle is to import contradiction into the intuitionistic logic which our comparison [with mathematics] makes the natural choice for practical judgements.¹¹²

So, Wiggins' challenge is to show why we should accept that moral discourse is truth-apt in the first place, given the actuality of absolute moral un-settleability.¹¹³ Wiggins bases his claim, as I interpret it, that there are absolutely un-settleable practical questions on cases of irresolvable personal dilemma – on the existence of hard, underdetermined, choices of conscience. In the next chapter, I shall argue that there is no reason to see the – real – phenomenon of moral dilemmas as a reason to abandon moral truth in the way Wiggins argues here.

3) But, there is a third possibility – the option, I argue the minimalist should adopt. We can argue for the rejection of the claim that Wiggins just took for granted: That we know that there are absolutely un-settleable moral disputes. In fact, given our investigation of the idea of absolute moral un-settleability, we have come to see this claim as un-justifiable. How can we ever justify a priori the claim that, for all future, we will never find an objective basis in either of the ways discovered on which to solve all radical disputes? In particular, as we have seen, this would involve showing a priori that moral discourse does not supervene on any other objective discourse, that no intuitional epistemology can ever be devised for morals and that moral truth is evidentially constrained. This, in turn, would involve showing a priori that none of the great projects of meta-ethics – for instance, naturalism, intuitionism, Kantianism and divine command theories – will ever succeed. And this, surely, looks like a case of a philosopher imprudently making claims about the impossibility of inventions being made some time in future – in particular, the invention of some completely novel way of deciding moral matters.

¹¹² Wiggins (1998), p. 130-31.

¹¹³ Wiggins (1998), p. 177. Incidentally, Wiggins goes on to develop his own position by claiming that whereas practical discourse, thus, does not admit of 'plain truth' – since it does not satisfy the convergence commitment – we can still accept a weaker form of cognitivism: it still makes sense to discuss, criticise and advice on moral issues and lack of guarantee of convergence is no reason for anyone to stop searching for *the right answer*. The distinction between a weaker and a stronger form of cognitivism is akin to the minimalist's distinction between truth-apt discourse that satisfies Cognitive Command and discourse that doesn't. However, the difference between Wright's minimalist and Wiggins' under-determinationist – weak – cognitivist is that the former works with a genuine notion of truth, whereas the latter rejects 'plain truth'.

So, there appears to be a good reason to refuse to accept that we have evidence that there are absolutely un-settleable moral disputes. Thus, we have independent good reason to avoid *Scylla*. Hence, the minimalist is not to worry about the inconsistency that *Scylla* would get him into.

6. *Second proposal for characterising actual moral disagreement – Charybdis.*

A second proposal for characterising actual moral disagreement is to rule out a priori that any of the de facto un-settleable disputes could turn out to be absolutely un-settleable. To go to the other extreme, so to speak. That is, claim that this principle is true:

P3 It is a priori that all moral disputes are settleable.

This is, of course, equivalent to the claim that it is *a priori* that any moral dispute must involve a cognitive shortcoming – i.e. that moral discourse exhibits Cognitive Command. If this claim is true, then epistemic constraints on moral truth is consistent with the a priori acceptance of bivalence for moral discourse. This kind of moral objectivism is clearly opposed to minimalism – that involves rejecting that moral discourse exhibits Cognitive Command – and the minimalist must avoid it like Charybdis.¹¹⁴ However, why is it unattractive in the first place?

We may choose to view the moral objectivist as a very strong optimist about the prospects of settleability and determinacy in ethical discourse – the kind of optimist that believes he has an *a priori* proof of success. Before arguing that this very strong kind of optimism is untenable, it should be distinguished from two weaker kinds of optimism that are more acceptable.

The first type of weak optimist about moral settleability is the one that believes that we can accumulate *a posteriori* evidence for the claim that all moral disputes are settleable – that as a matter of fact people are converging more and more in their moral beliefs due to the fact that methods for resolving moral disputes are evolving. A view like this has been defended for moral discourse by philosophers like

¹¹⁴ The character, I am here calling a moral objectivist, thus, is distinguished by his making three claims: That moral truth is epistemically constrained and that both universal settleability and the principle of bivalence holds true a priori in moral discourse.

Michael Smith and Derek Parfit. They argue that moral debate often in the past has been influenced by irrational authoritarianism and arbitrary religious belief. They argue that is not irrational to hope that as these elements are gradually being rooted out of human discourse – as it becomes more free of coercion and, hence, more rational – we will eventually achieve a rationally established agreement. As Parfit says: "We cannot yet predict whether, as in Mathematics, we will all reach agreement [in ethical discourse]. Since we cannot know how Ethics will develop, it is not irrational to have high hopes".¹¹⁵ This form of '*a posteriori* based optimism' seems very plausible and can be endorsed by the minimalist.

The second form of weak optimism is merely pragmatic. This is the thought that, since the whole indispensable practice of moral argument and criticism hinges on belief in the possibility of reaching convergence in moral verdicts, we had better act *as if* we knew that convergence can be achieved. If not, we risk losing a valuable tool for creating a moral consensus – and, hence, the conditions for social co-operation and dynamism in our moral belief-system. For pragmatic reasons, therefore, we should always persevere in the face of lack of settlement and determinacy – including lack of *a priori* or *a posteriori* reasons for believing in moral convergence under favourable conditions. Again, the minimalist cannot object to this form of optimism.

Contrasting the objectivist's – the believer in P3 – optimism with these two more reasonable degrees of optimism, it is clear that his is much more difficult to defend. Let me now ask, what it would take to defend strong objectivism. What does it take to provide an *a priori* guarantee of universal settleability in ethics? There would seem to be three ways of showing that moral discourse exhibits Cognitive Command:

First of all, we might show that the principle of bivalence is guaranteed to hold in morals. To see that a guarantee of bivalence in moral discourse will give us an *a priori* case for settleability in ethics, notice that Shapiro and Taschek's argument almost gave us a proof of universal settleability. It *did* give us:

It is *a priori* that, for any dispute, it is not the case that there are no cognitive shortcomings on either side of the dispute.

¹¹⁵ Parfit (1984), p. 454.

But, if – assuming classical logic – we get rid of the double-negation in this result, we have a statement of a guarantee of universal settleability. But, the double-negation-elimination step here, of course, begs the question against the minimalist. If we must accept that we have no *a priori* guarantee of settleability in moral discourse (as the minimalist contends), then, given an acceptance of epistemic constraints on moral truth, we must also lose any guarantee that bivalence holds universally in moral discourse – we can no longer assume that we have an *a priori* warrant for bivalence in moral discourse. But, then we must likewise abandon the double-negation elimination-rule in this context. Thus, the prospect of proving that settleability is guaranteed *a priori* in this way presupposes that settleability is guaranteed *a priori*. The minimalist cannot be rebutted in this way without assuming that he has already been rebutted. (Of course, the objectivist cannot be rebutted without circularity either).

Another way of attempting to show that moral disputes are guaranteed to be settleable, is by showing that actual counterexamples do not exist and cannot even be imagined. Thus, we cannot imagine counter-examples to universal settleability with regard to simple arithmetical truths, let alone produce any. However, it seems that counterexamples to the claim about a guarantee of a universal moral settleability are, at least *prima facie*, imaginable, as the many possible examples investigated in section 3 clearly indicate.

So, we are left with a third suggestion for getting an *a priori* guarantee of settleability. We might prove P3 by simply providing – via conceptual analysis – an *effective method for deciding all moral matters*. The way to show this would be to show that competent participants in the discourse are in possession of an effective procedure for generating, in every case, a compelling (though defeasible) case for or against a given statement.

Arguably, examples of discourses for which such a proof could successfully be constructed are discourse about simple arithmetic and discourse about the colour of objects in the immediate surroundings. Both discourses for which we can plausibly devise an intuitional epistemology. Here, plausibly, I – as a competent user of the relevant concepts – do have a procedure for examining a question such that, if I carry it out correctly and appreciate the outcome correctly, I am *bound* to be correct. And if it turns out that I am not correct, I will know that at least one of the assumptions I made about having carried out the procedure and appreciated the result correctly is

false. For example, it might be false that I didn't forget to carry when I added or that I wasn't wearing tinted spectacles when I made my colour-observations. Likewise, it will *always* be the case that, if two people carry out the procedure, yet get divergent result, then at least one of them will have made a cognitive mistake – at least one of the assumptions made in the course of carrying out the procedure must be false. These cases, I suggest, may be seen as paradigms of discourses that satisfy Cognitive Command due to their intuitional epistemologies.¹¹⁶

But, what would such a procedure look like in ethics and do we have any? Well, as we saw when we defined the notion of settleability, we can find an objective basis for a resolution of a moral dispute in several ways. We can sometimes settle a moral dispute by finding false non-moral or moral background beliefs, or flawed reasoning, on the part of one or both parties to the dispute. But, as was also argued, this form of settlement cannot possibly be achieved in all moral disputes, since it is a feature of moral practice that a moral dispute may be over the acceptance of a *basic* moral linking-principle – a belief not inferentially supported by further background evidence and inference-principles, but to which we may have an affectively grounded entitlement.

So, what *could* justify the objectivist's staunch optimism about the settleability of radical moral disputes? It seems that what has to be proved is that the possession of an effective decision-method for moral disputes is built into the very competence with moral concepts.¹¹⁷ By analogy to the paradigm cases of decidable discourse – arithmetic and colour-perception – this amounts to the task of showing that the adoption of basic moral principles is answerable to a special moral intuitional faculty. Like basic truths of arithmetic and colour-perception, on this view it would be possible to (defeasibly) demonstrate the truth of basic moral principles through carrying out a certain procedure (inherent in moral concepts) such that it is guaranteed that if all the assumptions made while carrying out the procedure are true, then the outcome of carrying out the procedure is bound to be true.

¹¹⁶ The account of decidability that I give in this paragraph is inspired by Wright's notion of 'actual verifiability'. See Wright (1993), p. 117-120.

¹¹⁷ Notice, that one way to achieve this kind of guarantee of cognitively blameless disputes in morals is to show *a priori* that moral discourse disputationally supervenes on some other discourse in which matters are guaranteed to be determinate. I shall consider the – bleak – prospects of this road to objectivity in morals – i.e. some form of reductionist naturalism – in chapter 5.

Hare is an example of a philosopher who believes that something like the decision-procedure required for constructing an *a priori* case for universal settleability is inherent in the very meanings of moral terms of our shared moral language:

'Once we see that there can be a common language, containing words like 'right' and 'wrong' and 'ought', which people of quite different moral opinions, from quite different cultures, can share, we can also see that they may be able to discuss their moral differences with one another in this same language. The question about objectivity then turns into the following: are there any logical properties of this common language, any rules governing its use, which will constrain those who think rationally about these questions to come to the same opinions, even if they start with different ones? The answer... is that there is'.¹¹⁸

Hare goes on to claim that a feature of moral concepts perhaps first noticed by Kant – that they require the universalisability of moral claims – *will* constrain rational moral deliberators to the extent that convergence is guaranteed. The other great projects of meta-ethical philosophy – naturalism, intuitionism, divine command theory – set out to provide a similar *a priori* guarantee. The minimalist can agree with much of what Hare here claims. In particular, that what it takes for the objectivist to defend his optimism is to justify *a priori* – through an analysis of the logical features of moral terms – a set of decision-procedures for moral judgements. However, Hare's last affirmation is where the minimalist chooses to remain agnostic.

So, what is the minimalist's reason for being less optimistic about moral discourse than the objectivist – to avoid Charybdis? Well, he simply does not believe that any of the projects attempting to show that practice with moral concepts necessarily involves the possession of effective procedures for decision of all moral questions have been successful. The 'logical properties' of moral terms – the core platitudes – appear, *pace Hare*, not to contain the means to compel all rational investigators to converge. In particular, the above-mentioned platitudes about moral debate appear to provide only a coherence-epistemology (though perhaps with an affective grounding) for discourse about moral principles. And, crucially, a demand for coherence is not in itself sufficient to guarantee convergence. We have no *a priori* guarantee there are not equally coherent, but incompatible, moral systems. Charybdis is the mistake of assuming that we have such a guarantee.

¹¹⁸ Hare (1999), p. 65.

On the other hand, to re-iterate, this is not to rule out completely the possibility that such a guarantee could not some day be found – that the successful completion of one of the projects of meta-ethics can never 'take up the slack' in moral reasoning, to use Wiggins' formulation, such that we will one day have such an *a priori* case for settleability. To think that the possibility of such progress in moral philosophy is not at least a possibility would, again, amount to being ensnared in Scylla!¹¹⁹

7. Minimalism and moral disagreement.

To sum up: the minimalist attempts to steer a course between two evils. He attempts to avoid *Scylla* – i.e. claiming of specific disputes that they are absolute un-settleable and, thereby, giving up too early on the prospect of a resolution. But he also attempts to avoid *Charybdis*, namely the error of leaping to the assumption that we have an *a priori* guarantee that all ethical questions are settleable. In his response to the problem of characterising actual moral disagreement, the minimalist manages to steer between Scylla and Charybdis by claiming that we should remain agnostic about the truth of any *a priori* claim that moral disputes are settleable.

P4 It is not guaranteed *a priori* that all moral disputes are settleable.

The minimalist, in other words, recommends that we remain agnostic about *a priori* claims about the completeness of moral reasoning – we must grant that we have, so far at least, no *a priori* grounds for holding that, for all moral *p*, either *p* is knowable or not-*p* is knowable.

As the foregoing has shown, the claim that we should steer between the two poles amounts to what might be called *a cautious optimism* about the potential for

¹¹⁹ Here I am, I think, supported by Wiggins (1998), p. 178-179. 'One sort of realist, a utilitarian, for instance, or a Neo-Kantian or some other sort of foundationalist, might say that the only reason why underdetermination has been allowed to win through to this point is that I have ignored the possibility that *moral philosophy itself*, e.g. in the shape of his (the Kantian's, the Utilitarian's or some other) theory, might take up the slack that prevents the convergence-condition from being satisfied; and that, when a more perfect moral philosophy emerges, human beings will then begin to converge in their moral beliefs, and will converge in their judgements precisely because there will be nothing else to think [Wiggins refers here to Hare]. To this I should respond that there is this bare possibility; but as yet no position has shown signs of mustering even one distinctive, substantial premise that will possess the second mark of truth'. (The second mark of truth speaks of 'commanding convergence').

progress towards a rationally imposed convergence in moral discourse. On the one hand, it cautiously expresses a scepticism about the possibility for a systematic practical rationality, that is, a rationality that gives us decision-procedures with the ability to give us determinate answers to all moral questions. The ability to *always* transcend the *mores* that initially animates people to engage in particular moral disputes – their affectively grounded set of basic moral beliefs – and compel every rational disputant to reach the same verdict in the end. This caution is, I believe, also expressed by Wiggins (given that he does not hold true what follows the two occurrences of 'unless'):

‘Of course, unless a fully formed universal notion of rationality exists in advance of the attempt to surmount [the human sensibilities that arises from practices at particular times and places] – unless there is a notion of practical rationality which is more than something immanent in actual norms and practices, more than Aristotelian – there is no guarantee (pace Kant) that morality can always or everywhere transcend its starting-point.’¹²⁰

On the other hand, the cautious optimist does not want foreclose the possibility of progress altogether. He concedes that there is a potential in our modes of moral reasoning and criticism to resolve conflicts and transcend starting-points. Or to consult Wiggins again:

‘Equally, though, no limit needs to be set in advance [to the process of transcending particular human sensibilities]. For the effect of morality's attaining a better understanding of itself.....is not to identify that clear limit but to make room for both possibilities, our ability to transcend in one case and our inability to transcend in another...The cognitivist need not make predictions, except to urge in general terms (and to try to show in particular cases) that, for the central core of morality, perseverance is the proper counsel.’¹²¹

As I go on in the following, these cautiously optimistic claims will become more spelled out and gain more support, I hope. The question is: Where exactly does the steering between Scylla and Charybdis take the minimalist? In particular, I shall ask what the minimalist can say about a variety of questions and alternative meta-ethical positions: about the place of moral facts in nature, about the way the minimalist's

¹²⁰ Wiggins (1990-91), p. 76.

¹²¹ *op. cit.*, p. 76. For related doubts about the prospects of ethical theories, doubts about the prospects for philosophy providing "a general test for the correctness of basic ethical beliefs and principles" and an advocacy of a "theory about the nature of ethical thought that *leaves open* [his italics] the question whether there could be such tests", see B. Williams (1985), p. 72-74.

account of moral discourse fits with a naturalistic explanation of moral practice (hereunder in what sense minimalism can be classed as relativism), about what the minimalist can say about attempts to ensure moral objectivity through some form of naturalistic reduction of moral concepts or properties.

But, before this, I shall follow up on the problem for minimalism that, as I have already mentioned, Wiggins apparently points to. He suggests that the reality of moral dilemmas will force the minimalist to steer towards Scylla – and, hence, lethal inconsistency. I shall now attempt to show why this is not the case.

Chapter Four

Minimalism and Moral Dilemmas

In the previous chapter, I have argued that the seeming possibility (and perhaps actuality) of radical, interpersonal, moral disagreement does not pose a threat to the consistency of the minimalist's account of moral discourse. In particular, I have claimed that it does *not* tell against the minimal truth-aptness of moral statements and that it *does* tell against the objectivity of morality – when moral objectivity is conceived of as the *a priori* guarantee that all moral disputes involve a cognitive shortcoming.

In this chapter, I consider moral dilemmas. I will argue that the type of un-settleability that moral dilemmas give rise to is not the same as the one that I considered in the previous chapter. In particular, I will try to defend the claim that moral dilemmas are phenomena that challenge neither the truth-aptness *nor* the objectivity of moral discourse. Moral dilemmas – I shall argue – simply do not entail the appropriate kind of indeterminacy.

In the following, I will answer two different possible challenges to the minimalist position. One that I take to be posed by Bernard Williams claiming that the actuality of moral dilemmas shows that moral reasoning must be inconsistent – hence cannot be truth-apt. And one – deriving from a remark of David Wiggins' – that tries to show that moral dilemmas present us with absolute un-settleability which again under the assumption that moral truth is epistemically constrained gives us inconsistency. I shall argue that the answers to both of these seeming problems for the minimalist can be found when we adopt the proper characterisation of moral dilemmas (and the right logic for the notions used in this characterisation).

Let me first – before outlining the structure of this chapter – present the two challenges.

1. Two challenges.

The first challenge is found among other places in Bernard Williams' early writings¹²² on moral conflict and consistency. He famously argued that moral conflicts are more like conflicts of desires than conflicts of beliefs. Hence, he took the possibility of moral conflict and dilemma to speak for a 'non-realist' – that is, a non-cognitivist – interpretation of morality according to which moral statements are expressions of attitude rather than expressions of belief. The proposed argument – in my very short version – is as follows:

1. The set of correct moral judgements about a situation of moral dilemma or conflict is inconsistent.
2. No set of *correct* beliefs is inconsistent, whereas the set of correct attitudes can be inconsistent.

∴ Moral judgements express attitudes rather than beliefs.

The problematic claim here is his premise 1. Consider the following quote:

‘...the non-realist approach may well allow for the possibility that one can be forced to two inconsistent moral judgement about the same situation, each of which backed by the best possible reasons, and each of them firmly demanding acceptance; and while action or advice demands deciding between them, it does not demand – or permit – deciding that either of them was wrong, or only apparently a requirement of the situation.’¹²³

Now, there is much in this characterisation that will survive my criticism in the following, but crucially I will want to argue that whereas there is of course a need to recognise the possibility of conflicting ought-statements in conflicts and in dilemmatic situations, there is no need to speak of *inconsistency*, and hence no need to give up moral cognitivism.

The second challenge is in essence the same as the first, but it appears to have an additional edge towards the minimalist, since it seems to be working under the minimalist's assumption that moral truth is epistemically constrained. Wiggins assumes that there actually are tragic moral dilemmas:

¹²² Williams (1965) and (1966).

¹²³ Williams (1966), p. 205.

'It seems that in the sphere of the practical we may know for certain that there exist absolutely undecidable questions – e.g. cases where the situation is so calamitous or the choices so insupportable that nothing could count as *the* morally reasonable answer.'¹²⁴

I interpret this as the claim that *moral dilemmas* are real examples of absolutely undecidable or un-settleable moral questions. He draws the conclusion that this real phenomenon ('knowable with certainty') presents the cognitivist (even of the anti-realist sort who demurs with respect to the principle of bivalence) with devastating inconsistency:

'If we insist upon the actuality of some absolute undecidability in the practical sphere, then we shall burst the bounds of ordinary, plain truth. To *negate* the law of excluded middle is to import contradiction into the intuitionistic logic which our comparison [with mathematics] makes the natural choice for practical judgements. The *denial* of '(A would be right) or not (A would be right))' contradicts the intuitionist theorem '(not(not(p or not p)))'.¹²⁵

Thus, again the truth-aptness of moral discourse appears to be threatened by the phenomenon of moral dilemmas. I take the argument spelt out in full to be something like this:

1. If there are absolutely un-settleable moral disputes and if moral truth is epistemically constrained then moral cognitivism is inconsistent.
 2. Moral dilemmas are examples of absolutely un-settleable disputes.
 3. There are moral dilemmas.
 4. Moral truth is epistemically constrained.
-
- ∴ Moral cognitivism is inconsistent (hence false).

The first premise here can be established by the following argument – showing that epistemic constraint and absolute un-settleability is inconsistent (even for the logical intuitionist):

1	(1)	$\diamond Kp \leftrightarrow p / \diamond K\neg p \leftrightarrow \neg p$	EC
2	(2)	$\neg(\diamond Kp \vee \diamond K\neg p)$	Moral dilemmas (Incompleteness)
1,2	(3)	$\neg(p \vee \neg p)$	1,2, substitution ($p/\diamond Kp, \neg p/\diamond K\neg p$)
	(4)	$\neg\neg(p \vee \neg p)$	Intuitionistic (& classical) theorem
1,2	(5)	\perp	3,4 $\neg E$
	(6)	\neg (EC & Dilemmas)	

¹²⁴ Wiggins (1998), p. 130. See also, p.174-6.

¹²⁵ Wiggins (1998), p. 130-31.

My answer to this challenge, to anticipate, will be that although there is a sense in which moral dilemmas present us with a matter that cannot be settled by moral reasoning – namely *which* act to perform – it does not present us with moral unsettleability (or undecidability or incompleteness) in the *relevant* sense. That is, in the sense that it presents us with moral judgements – moral ought-statements – whose truth-value we cannot even in principle determine. Hence, premise 2 is false.

To meet these challenges, the minimalist must show that the best account of moral dilemmas is compatible with his cognitivist, non-objectivist and anti-realist account of moral discourse. First, however, I want to become clearer about what makes moral dilemmas relevantly different from the cases of moral disagreement considered in the previous chapter (section 2). And I want to distinguish two different senses of inconsistency (section 3). The crucial task, I shall argue, is then to characterise moral dilemmas in the proper way (sections 4 to 7) – hereunder to account for the logic of moral dilemmas (section 8). Finally, I am then ready to answer Williams (section 9) and Wiggins (sections 10 and 11).

2. Moral conflict and moral disagreement.

Different people at the same time and the same person at different times can *disagree* over some moral issue. One way to point to the distinguishing feature of a moral *dilemma* is to notice that only *one* person at *one* time can experience it. It is a kind of moral disagreement or conflict with oneself at a particular point in time in a particular set of circumstances. I will want to argue that this difference is very important. In particular that, on reflection, moral dilemmas do not present us with un-settleability in the sense that radical moral disagreement did.

As we defined the term in the previous chapter, there are three relevant possible sources of moral un-settleability. Moral disagreement, we saw, can derive from: 1) disagreement over the adoption of a single principle or value (as a basic principle or value), 2) disagreement over the extension of a single moral predicate, and 3) disagreement over what principles of ranking or weighing to use when trying to reach a practical verdict – an answer to the question about what one ought to do,

all things considered – in a particular situation in which *different* values or principles are deemed relevant.

It appears to me that neither of these three possible sources of un-settleability is relevant to intra-personal conflict. To see this, imagine first what it would be like to disagree with oneself (at a particular point in time) over the adoption of a single moral principle. One would have to adopt two principles that were strongly inconsistent – in the sense that there is no possible world in which one could consistently judge in accordance with both principles. For example, one might adopt the principles that intentional killing is *always* morally wrong and that euthanasia is morally permissible. Clearly this would just be a case of holding inconsistent beliefs and that will show us nothing more interesting about the truth-aptness or the objective status of moral discourse than the mere possibility of holding inconsistent scientific beliefs tells us about the objective status of science.¹²⁶ In any case, this is not what we would call ‘being in a moral dilemma’.

Also, it seems that I cannot disagree with myself over where to draw the boundary of the extension of a single moral concept. *I* will know where to draw the boundary (though perhaps I may find it a hard case). Problems only arrive if *someone else* (who is well-informed and competent) comes up with a different way of drawing the boundary. Compare here with the way semantic vagueness in a colour-predicate can cause dispute over a borderline case. *I* can easily settle on a particular colour-verdict – I just have to report how the sample strikes me. That's why I enter into a dispute with someone else in the first place. There is nothing resembling an *inner* conflict or dilemma here.

We are left with the claim that moral dilemmas are cases of lack of settleability down to an inner conflict between (meta-)principles for weighing or ranking moral considerations in the context of making a particular practical verdict. But, the idea that moral dilemmas derive from internal disagreement over what meta-principle to adopt seems to have the same problems as the case of disagreeing over the adoption of a single principle. It implies that one adopts contradictory principles at one and the same time – for example something of the general form: 'Moral considerations of type

¹²⁶ A point made also by Williams (1965), p. 171.

X outweighs moral considerations of type Y in circumstances C', and its negation. And, again, that is being inconsistent, not being in a moral dilemma.

3. *Moral conflict and inconsistency.*

So, we have already some reason to believe that the un-settleability involved in moral dilemmas is not of the same type as the un-settleability involved in radical moral disagreement. Before turning to the task of describing what type of un-settleability *is* involved in moral dilemmas – how to characterise moral dilemmas – I want to make some observations about moral conflict (and the special case of moral dilemma) and moral inconsistency in order to clarify what Williams meant when he challenged the cognitivist to account for the apparent inconsistency involved in moral reasoning. Notice first of all that Williams argues that moral inconsistency follows not only from moral dilemmas, but from the existence of any moral conflict – even, that is, cases where one moral reason clearly outweighs or overrides some conflicting moral reason.

Williams does not want to claim that moral conflict causes inconsistency in the strong uninteresting sense mentioned above – cases where two principles are held true that cannot possibly be true together. What he is thinking of are cases of *weak* inconsistency. That is, a contingent inconsistency in one's set of moral commitments – an inconsistency that only crops up given certain ways the world happens to be, under certain empirical conditions. Let me give a down to earth example of weak inconsistency taken from Blackburn.¹²⁷ I may happily have let myself be guided for many years by two (non-ranked) principles when shopping: 'Never buy the cheapest brand' and 'Always follow your mother's recommendations on what to buy'. However, when one day mother recommends the cheapest brand, my shopping-code shows itself as weakly inconsistent and breaks down. But, the breakdown was merely contingent on that fateful recommendation of my mother. Had things turned out differently, the code might have survived.

Now, of course even when we recognise what Williams' claim about moral inconsistency really amounts to, there is still a threat to minimalism. Weak inconsistency is inconsistency none the less. But, even though I will ultimately dismiss

¹²⁷ Blackburn (1996), p. 130-31.

this claim of Williams', there is perhaps still an important insight in Williams' proposal. What he is trying to do is to capture an important feature of moral discourse – the fact that moral reasons frequently conflict and that a rational and entirely happy moral resolution sometimes in some sense is unachievable. There is a clear sense in which a moral reason stays with us, even when it is overridden or outweighed by stronger moral reasons. And what is more, this feature of morality is a sign of maturity of moral reasoning rather than – as it would be in cases of reasoning about scientific matters – a sign of unfinished business. The problem with Williams' way of capturing this phenomenon is only that he chooses to characterise it in a way that entails inconsistency.

To see why it is important in one's account of moral conflict and dilemma to accommodate the irreducible phenomenon of moral conflict rather than to try to prevent it from ever occurring, consider that it is relatively easy to diminish the probability of these moral conflicts occurring. One can do this, for instance, by advancing a moral theory that limits the principles and values to an absolute minimum – for instance, the value of well-being and the principle 'One ought to maximise good in the world'. I would here ensure that contingent conflicting verdicts on what to do, moral conflicts, will be a much rarer phenomenon in the same way as I could prevent – perhaps influenced by Buddhism – the frustration of my desires by cultivating an indifference to the outside world. Or, to take another example, one can avoid many conflicts of obligations simply by never promising anything and by never forming any close personal relationships or never having any children.

However, as the analogy with desires already may have suggested, it is not clear why we should see trying to avoid conflict at any cost as a worthy goal for our moral thinking. Just as we may think that avoiding frustration of desires by attempting not to have worldly desires at all would result in an impoverished life, so the attempt to insulate one's moral beliefs from conflict and dilemma by simplifying them, or by trying to prevent conflict-creating contingent circumstances to occur, would seem to merely cause one to become insensitive to the complexity of moral considerations that a mature moral agent would recognise.¹²⁸

¹²⁸ This analogy is inspired by Williams (1965). I am of course here voicing my own (moral) opinion – a belief in a pluralistic and incommensurabilist account of morals.

I should mention that there is a way in which avoiding conflict and dilemma in our moral theory may be thought not to involve insensitivity and simplification. That is, one might make a heavy metaphysical assumption about there being a certain moral order in the universe. Perhaps God has pre-established a moral harmony in the world such that agents who have the right moral beliefs (maybe highly pluralistic and sensitive) and have acted rightly in the past will never experience moral dilemmas or important moral conflicts. Thus, there would always be a definite answer to the question about *the* right way to act for the virtuous person. If this were the case, then conflict and dilemma would be a mark of inadequate moral thinking or previous wrongdoing and their occurrence would be a reason to change one's moral beliefs or feel guilty about one's past actions.

However, until this very strong metaphysical assumption is argued for, it would seem that moral conflict and dilemma are irreducible phenomena that anyone thinking about ethics should somehow recognise. But, as I shall try to argue *pace* Williams in the following, there is a way of characterising and capturing these phenomena that does not have us saying that inconsistency is a necessary feature of moral reasoning.

4. Towards a characterisation of moral dilemmas

Not every inner moral conflict is a true moral dilemma. There are cases where inner moral conflicts – understood as struggles between moral considerations pro and con in a particular situation – are rationally resolvable. That is, cases where one set of moral considerations overrides or outweighs the other and the merely *apparent* inner dilemma is settled. Moral dilemmas should, instead, be seen as a subset of moral conflicts. They are – in some sense to be specified in the following – conflicts of moral considerations that are not settleable.

However, there is reason to be very careful when characterising moral dilemmas – when deciding how exactly to circumscribe the relevant class of un-settleable conflicts. It would seem that different technical definitions are given depending mainly on what further philosophical pursuits the definer is engaged in. I will want in the following to define 'moral dilemma' *in a way that makes it relevant to our discussion of moral truth and objectivity*. And this pursuit will cause me to leave

out of my definition reference to what is thought by many to be essential features of moral dilemmas. My definition will be broader – will include a wider range phenomena.

I will arrive at a definition after working my way through a set of examples of dilemmas and after discussing and at least partially modifying a proposed definition by Walter Sinnott-Armstrong. Let me, therefore, get started by citing Sinnott-Armstrong's influential definition of a moral dilemma and then use this as raw-material for refinement in the following.

Any situation is a *moral dilemma* where at the same time:

- 1) There is a moral requirement for an agent to adopt each of two alternatives.
- 2) Neither moral requirement is overridden in any morally relevant way.
- 3) The agent cannot adopt both alternatives together
- 4) The agent can adopt each alternative separately.

This should be fairly self-explanatory, though I will mention that the reason why the fourth condition is included is that, if it were not, then pairs of totally unrelated moral non-overridden requirements could constitute a moral dilemma if only one of the requirements were impossible to fulfil. For in that case, conditions 1), 2) and 3) would all be satisfied. The last of these, since I cannot adopt both alternatives simultaneously, if I cannot adopt one of them.

I want to focus on and discuss two questions that Sinnott-Armstrong does not address or – to my opinion – addresses in the wrong way in his discussion of his own definition. Firstly, I want to argue that the existence of moral dilemmas must be accepted whether one favours a teleological or a deontological theory of ethics. Thus, our main concern here – getting clear about the implications of the possibility of moral dilemmas for moral truth and objectivity – is left untouched by certain discussions in normative ethics. Secondly, I want to briefly argue that we should not (*pace* some in the debates) want to restrict the term 'moral dilemma' to be used in situations where an agent is bound to experience a strong form of moral residue – for instance guilt or remorse – no matter how he practically resolves the dilemma. Thus, I will briefly address the issue of what the appropriate moral residue after the resolution of a moral dilemma could be.

However, the main contention in the following will be that we ought to use the notion of 'a non-overridden moral requirement' to characterise moral dilemmas as opposed to 'an overriding moral requirement' or an 'all things considered, moral requirement'. In this, I agree fully with Sinnott-Armstrong, as shall be evident.

5. Dilemmas, deontology and teleology.

To set the stage and fix the terminology before going into the first question – about the neutrality of normative questions about deontology versus teleology relative to the issue about moral dilemmas – let me digress to give a brief sketch of these two different conceptions of moral practical reasoning. We may say that through practical reasoning, we try to determine what, all things considered, we ought to do in particular situations. We engage in such reasoning to be guided morally in our choices of action. In doing this, we try to bring our values to bear on our conduct. The agent will have a set of values that are relevant to a given situation, where a choice of action is called for. But, how do general values turn into reasons for action in particular situations? There would seem to be two general ways of effecting this transformation – the deontological and the teleological (or consequentialist).

Consider a simple (tediously well-known) example of a moral conflict where only one value is in play. Tom is faced with a tragic choice. He has just noticed that a trolley is speeding down a track towards a group of 3 people working on the track. Tom is situated in a distant rail control-tower where he works and, unfortunately, from there he cannot alert the group in time. However, he knows that by pulling a lever, he can divert the trolley onto another track. But, he also notices that one other person is standing on this second track. Tom has a few seconds to decide. Should he pull the lever to save three lives, or should he choose not to interfere to avoid intentionally causing the death of the person on the diversion-track?

Now we can imagine two moral deliberators, call them Immanuel and John. They agree that human life is very valuable and they agree completely in their non-moral description of the situation, but they still end up giving contrary verdicts on what Tom ought, all things considered, to do. Their difference of opinion here ultimately derives from their differences over *how* one ought to let one's values determine what one ought to do.

Immanuel believes that our valuing life should make us show *respect* for or *honour* life. He is a deontologist. That is, on every occasion when we act, we should bear in mind that our action should not directly cause death. We should not allow ourselves to become the instrument of evil. Therefore Immanuel has adopted a moral norm: Deliberately killing the innocent is wrong. This norm forbids directly actions of a certain description, namely a type of action showing a lack of respect for life. What further consequences this action may have is irrelevant. Hence, his advice to Tom would be: You ought not to pull the lever.

John, on the other hand, believes that we should value life by *promoting* it. He is a teleologist or a consequentialist. That is, we should, according to him, always act in the way that maximises the amount of good in the world – that is, in the way that has the best consequences.¹²⁹ When deciding, we should, first of all, line up the available alternatives for action. Then we should predict what the different likely consequences (outcomes) of these available alternatives are, including assigning probabilities to the outcomes. On the basis of this, we should assess/calculate how much good we would expect to bring into the world when choosing the various available alternatives for action. Finally, we should single out the best possible action among the alternatives – the action (or there may be more than one: the actions) that is *expected* to have the best overall consequences. Applying this procedure to Tom's tragic circumstances, it is clear to John that his advice to Tom must be: You ought to pull the lever.

Now, John and Immanuel could have been different voices in the head of Tom. And I want to say that a normal agent – when summoning all relevant considerations in order to determine what he ought to do (all things considered) – will tend to come up with considerations of both the teleological and the deontological type.¹³⁰ Thus, it is not (strongly) inconsistent for Tom to hold both the principle that

¹²⁹ Or in a lesser known, *satisficing*, version of consequentialism, we should choose the action(s) that bring(s) about a certain level of good in the world.

¹³⁰ It should be mentioned that both deontologists and teleologists can try to assimilate the type of reasons associated with the other theorist into their preferred moral reasoning. The teleologist can work with a pluralistic theory of values that include things like promise-keeping, honesty and equality – he can see these things as goods in themselves. He can then allow the instantiation of these values to weigh very heavily in his assessment of the various consequences of actions – hence reaching much the same actual practical verdicts as the deontologist. On the other hand, the deontologist can include imperfect duties – duties to help others, and to develop our own talents, to take examples from Kant – in his list of duties. Imperfect duties are duties that need not be carried

he ought to *promote* life by seeking to do the thing that maximises life and at the same time the principle that he should *respect* life by never killing the innocent intentionally.

In many cases, these different principles will work together to bring about a clear verdict on the question what he ought to do. In these (uninteresting) cases, we can say that the considerations are *harmonic*. In other cases, the considerations work against each other, yet in such a way that one set of moral considerations clearly overrides or outweighs the other. These will be cases of (*resolvable*) *moral conflict*. But, in Tom's tragic case it might seem that the considerations both count against one another and seem to leave no room for deciding which consideration would win. His case we will therefore call a *moral dilemma* or an unresolvable moral conflict (realising of course that this definition will have to be refined considerably in the following).

Notice now, that a great variety of conflicts and dilemmas are possible given the above account of moral practical reasoning. Even assuming that only one value is in play (as in the above case), we can conceive of three general types: 1) Conflict between consideration deriving from deontological principles – considerations of respect for values. Think of a symmetric case. I value promise-keeping. I have, however, made promises to two different persons that (initially unbeknownst to me) I cannot both keep. 2) Conflict between considerations deriving from teleological principles – considerations about the promotion of a value. Think again of a symmetric case. The doctor has only one dose of morphine left, but two patients in severe pain justifying the use of morphine. Here, moral considerations speak for both of two incompatible actions (wherefore there are considerations speaking for and against each action) – each promoting the value of well-being. 3) Conflicts between considerations deriving from both principles of promoting and respecting a single value. We have already considered an example of this – Tom's dilemma.

If we assume that *more than one value* is in play, then the possibilities of conflict multiply. In particular, we get asymmetrical cases where considerations deriving from different values conflict and where, therefore, also the problem of incommensurability may emerge. Let us – without going into all the possible

out on every occasion, only so far as possible. This will effectively play the role of promoting certain values in the deontologist's moral reasoning. In short, the more you develop the theories of the teleologist and the deontologist, the closer they come together.

combinations – see how the teleologist and the deontologist would conceptualise this phenomenon of multi-value conflicts.

We may first think of a case where one is called to *promote* different values in the same situation. To take here an example of prudential practical deliberation, think of someone having to decide whether to accept a new job-offer. Speaking for taking the job is the consideration that, it would allow him to continue his research and thereby promote the value of understanding. Speaking against taking the job is the consideration that it will have costs in his private life due to his having to spend more hours on the job. In other words, it would mean not promoting (relative to not choosing the job) the value of deep personal relations.

Notice here also, that this may (but need not) be seen as a case of conflict of incommensurable values. In this case, there would appear to be no answer to the question of what one ought to do all things considered, since there would be no way of comparing the value of the two available courses of actions – no way to reach the conclusion that one set of considerations *outweigh* the other.

We may also imagine cases of conflict deriving from the attempt to *respect* different values in one situation. Think here of a simple case where you have made a promise to someone, but things turn out so that the only way you can keep this promise is by telling a lie to someone else. Here two deontological principles: 'Lying is wrong' and 'Promises must be kept' – adopted to respect the values of honesty and promise-keeping respectively – give rise to contrary considerations. This may then turn out to be a resolvable conflict – if the two obligations involved here can be *ranked* in this particular situation. However, there is also the possibility that the obligations cannot be ranked. In this case again there would appear to be no answer to the question of what one ought to do all things considered – no way to show that considerations for the action *overrides* considerations against.

To sum up these thoughts on the nature of moral practical reasoning, we seem to have discovered that moral conflicts can arise in cases where one wants to promote values, in cases where one wants to respect values and in mixed cases – in short, the phenomenon can haunt both John and Immanuel. Furthermore, we can expect that cases where all relevant moral considerations fail to give a conclusive practical verdict – where due to symmetry or incommensurability neither the pros nor the cons of some action overrides or outweighs the other – will occur both when we reason

teleologically and deontologically. In other words, moral dilemmas can have many shapes and, it would seem, neither a deontological nor a teleological approach to moral deliberation will save us from the phenomenon.

6. *Tragic moral dilemmas, moral residue and un-settleability.*

A second point to make before settling on a definition of ‘moral dilemma’ concerns the appropriateness of moral residue – like emotions of regret, guilt or remorse – upon the practical resolution of a moral dilemma. Many in the debate have defined moral dilemmas in such a way that dilemmas necessarily become *tragic* dilemmas. That is, situations where the agent cannot act in a way that does not make it appropriate for him to feel a degree of remorse – as opposed to, say, merely regret. Sinnott-Armstrong is an example of this tendency. In order to argue that we should not define ‘moral dilemma’ restrictively in this way – given that we are here concerned solely with the possibility that moral dilemmas involve moral un-settlability – I want to have a look at his arguments for this point.

First of all, Sinnott-Armstrong insists that moral dilemmas should be defined as irresolvable conflicts between moral *requirements*. His definition of a moral requirement is :

‘There is a moral requirement to adopt an alternative if and only if it would be morally wrong not to adopt the alternative in a situation that is similar to the actual one in all relevant respects except that, if there is a moral justification for not adopting it in the actual situation, there is no moral justification for not adopting it in the otherwise similar situation.’¹³¹

Sinnott-Armstrong distinguishes moral requirements defined in this way from *moral ideals*. He brings out the difference between the two by comparing promise-breaking with failure to be charitable. If I fail to donate money to a charity on a particular occasion, I can still claim to be charitable, if I gave to three other charities yesterday.

¹³¹ Sinnott-Armstrong (1988), p. 12. Notice that the definition allows for requirements to disappear in circumstances where we *do* want to say that they disappear. Namely, conditions under which the requirement is *cancelled* – as when a promisee dies – or conditions under which the agent is *excused* – as when the agent is not responsible for the obtaining of circumstances that makes his fulfilling the requirement impossible. In these types of circumstances, not adopting the alternative is not wrong in

If I break a promise, on the other hand, it is no use pointing out that I kept three promises yesterday. That is, promise-keeping is a moral requirement in that *every* failure to live up to it is morally wrong in some sense. Charity, on the other hand, is a moral ideal in that failure to live up to it on a particular occasion may not be wrong.¹³²

Notice that the counterfactual in the above definition of a moral requirement ensures that a moral requirement for an alternative remains real or 'in force' even when overriding moral justification for an incompatible alternative justifies the omission of it. In other words, moral requirements do not disappear when they are overridden or matched. Therefore, the agent *has* to violate some moral requirement in a conflict between requirements, no matter which way out of the dilemma or conflict he chooses.

Now we can see that Sinnott-Armstrong's suggestion that we restrict ourselves to clashes between moral requirements when defining moral dilemmas in fact is an attempt to restrict the notion of a dilemma to tragic cases. That is, Sinnott-Armstrong believes that to violate a moral requirement always leaves a strong form of moral residue – in particular, that to have failed to live up to a moral requirement (as opposed to having failed to live up to a moral ideal in a particular case) makes it appropriate for an agent to feel remorse. Therefore, since one has to violate a requirement no matter what horn of a dilemma one chooses, and since reasons to feel remorse follows inevitably on any violation of a moral requirement, any agent facing a dilemma is in a sense doomed to have reasons to feel remorse. Thus, Sinnott-Armstrong's restriction to moral requirement conflicts in his definition of moral dilemmas ensures that he retains the tragic element that characterises what are often considered to be paradigm examples of moral dilemmas – i.e. the situations tormenting characters like Agamemnon, Antigone, Abraham and Sophie – as an essential part of moral dilemmas. The sense that no matter what the agent does when faced with a dilemma, he will do something that justifies remorse or even guilt.

Now, the point to make is that we should not want to accept Sinnott-Armstrong's definition in the following. First of all, and most importantly, his pursuit is different from ours. He is trying to capture the essential features of paradigm

a world similar in all *relevant* respects to the actual one except for moral justification – it would be wrong only if the nearby possible world also differed with respect to cancellations or excuses

examples of *tragic* moral dilemma. We, on the other hand, are out to capture what is relevant in the phenomenon of moral dilemma to the debate over moral truth and objectivity – hence we will want to focus on the apparent involvement of *unsettleability* in dilemmatic situations.

Secondly, though this is a side-issue and I shall not argue the point in full, it is not clear to me that we ought to accept that the proper resolution of *any* moral dilemma could render appropriate a strong form of moral residue. Sinnott-Armstrong and others, I hold, are simply not correct to assume that remorse or guilt is ever justified after the practical resolution of a moral dilemma – hence there cannot be moral dilemmas as defined by him.

The first point is fairly easy to argue. What *we* are interested in here are cases of rational un-settleability in moral reasoning. But those are cases where moral considerations or reasons – be they teleological or deontological in shape – conflict and where there is no way of deciding which side wins, no way of reaching a stable practical verdict of the form: All things considered, x ought to ϕ (where ϕ -ing is not a disjunction of two or more acts). And clearly there is no reason why these conflicting reasons should be particularly heavy or fateful – for instance such that acting against them gives reasons for feeling remorse. Even a case where very weak reasons conflict in an un-settleable way constitutes a relevant case of un-settleability. (Likewise, to return to Sinnott-Armstrong's example for illustrating the distinction between moral requirements and moral ideals, having equally strong reasons to give to two different charities can constitute a relevantly un-settleable problem).

The second point takes much more argument. First of all, it is essential to distinguish precisely between the two degrees of moral residue. Recognising that there is justification for *second degree moral residue* is recognising that things like offering explanations, apologies, expressions of regret (understood as the feeling that something bad has happened, rather than the state of wishing that one had acted in another way) and perhaps some compensation after the act are appropriate in the circumstances. These actions crucially do not imply that the agent is culpable – that there are reasons for the agent's feeling remorse or guilt. *First degree moral residue*, on the other hand, involves recognising that feelings of guilt and remorse are justified

¹³² This distinction – of Sinnott-Armstrong's – is a version of the traditional distinction between

and, therefore, it does imply that the agent did something wrong. The question is whether first degree moral residue is *ever* justified.

It does seem to be clear to me that second degree moral residue typically *is* justified when having acted against a moral reason as the outcome of a conflict or dilemma. This goes even for the situation where the moral reason acted against is clearly overridden by some other moral reason. Say that I had to break a trivial promise to meet for lunch to visit a friend who had been hospitalised with sudden, life-threatening, illness – with no chance of alerting the lunch-date in advance. To me it would seem entirely appropriate for someone to feel obliged, on meeting the person, he stood up, to explain what happened, to express regret and perhaps to compensate by inviting the person over for dinner at a later time. Why is this? I suppose it has to do with making sure that my actions – and in particular, my intentions – are interpreted correctly. That is, that I meant no disrespect to my lunch-date. That I took my obligation very seriously. I want to make sure that he believes the counterfactual that if I had not had the obligation to be elsewhere, I would have kept my promise to meet him. I want to establish that I had the best intentions and should not be mistaken for a promise-breaker. That there is no reason for blame or spite.

Notice again – to return to the previous point about the irrelevance of the debate about deontology versus teleology – that second degree moral residue is also sometimes appropriate when the agent is merely concerned with the promotion of good. To take an example, consider again the case of one dose of morphine and two deserving patients. Again, here it seems entirely appropriate for the agent to feel obliged to explain the situation to the unlucky patient, to express regret and perhaps to promise him that if more morphine will appear, he will be the first to receive it. Again, this residue is justified for the same reason as in the case of moral requirement conflicts. Second degree residues are ways of showing respect and good intentions in situations where one might easily otherwise be misinterpreted.

However, *first degree moral residue* is arguably appropriate only if one has acted against the recommendation of an all things considered moral verdict. In other words, it is appropriate for an agent to feel guilty and for others to blame him, only if

he has done something wrong – where ‘wrong’ is defined such that if the agent has acted against a moral reason *with moral justification* (i.e. justification deriving from a stronger, overriding, moral reason or an equally strong moral reason) then he has done nothing wrong. Hence, feeling guilty is *never* reasonable after the proper resolution of a moral dilemma or conflict. What, according to me, *is* appropriate (for everyone involved) is to feel regret – where regret is here conceived as the feeling properly following on the belief that something bad has happened. That in an ideal world this regrettable thing would not have happened.¹³³

I am not going to argue this point in greater detail. But, let me hint at two reasons for refusing to accept that a dilemma may involve the appropriateness of first degree moral residue no matter what one chooses. The first shows how Sinnott-Armstrong’s account counterintuitively will have to dissociate the appropriateness of remorse from the appropriateness of social criticism and perhaps punishment. The second speculates why we *sometimes* do have the intuition that choosing both alternatives render it appropriate for the agent to feel remorse.

Consider first Sophie’s famous dilemma in the novel *Sophie’s Choice*. Sophie is ordered by a Nazi camp-guard to choose one of her children to go to the gas-chamber. If she refuses to choose both will die, so she must choose one. Sinnott-Armstrong would hold that remorse is appropriate for Sophie, no matter what she does. But also, he admits that it certainly would be wrong to blame or criticise her for making a choice, let alone believe that she would deserve punishment. (At most we can criticise her if she does not exhibit remorse). I think it is impossible, however, to dissociate the kind of (justified) moral self-criticism which is (justified) remorse from (justified) external criticism. Does it make sense for someone to say: ‘I certainly don’t blame or criticise you for what you did, but of course it would be appropriate for you to feel some degree of remorse’? It seems to me simply not to make sense. So, since

¹³³ Sinnott-Armstrong, however, tries to make his claim good in another way. He argues that we have basic intuitions about the appropriateness of remorse following on the practical resolution of a dilemma. In particular, he claims that we intuitively think that, in comparison with a mere bystander, it is appropriate for the agent – i.e. the person who has to act in a way that violates a requirement – to ‘feel something more’. And this difference can only be explained as the appropriateness for the agent, but not the bystander, to feel remorse. I find this argument unconvincing. Why not say that the ‘something more’ is just *more* regret – the added regret of the agent that *he* had to do this terrible act, this bad thing? Incidentally, Sinnott-Armstrong does seem to recognise that remorse or guilt is too strong, even given this consideration. He toys with the

blame intuitively cannot be justified after the practical resolution of a dilemma, neither can remorse.

Secondly, I suspect that some of the cases which Sinnott-Armstrong uses to conjure up the intuitions about remorse being appropriate on either horn of a dilemma are special cases – cases where the agent brought the dilemma on himself by previous wrong-doing.¹³⁴ In these cases, of course, the agent may be to blame, no matter what horn of a dilemma he chooses, since he ought to have foreseen and prevented the dilemma from occurring. One example Sinnott-Armstrong gives is that of Juan. (The example is not strictly relevant, since it concerns the appropriateness of compensation rather than remorse. But it still serves to illustrate my point). "For example, suppose Juan promises to deliver two farmers two loads of vegetables to a market today, *but he waits too late*, so he cannot deliver both loads on time".¹³⁵ He then goes on to say that intuitively Juan ought to compensate the farmer whom he will have to disappoint. I have the same intuition in this case, but I think it derives from the italicised bit which seems to suggest that Juan brought the dilemma on himself. If the example was spelled out so as to indicate that Juan was not to blame for getting involved in the dilemma, I suspect that the intuition would fade.

In short, I suspect that first degree moral residue – for conceptual reasons – is *never* appropriate after the proper resolution of a moral conflict or dilemma. And second degree moral residue is *always*¹³⁶ appropriate after the resolution of both moral conflicts and dilemmas. Hence I suspect – without having given a conclusive argument – that there can be no dilemmas, if 'dilemma' is here defined such that dilemmas necessarily involve justified first degree moral residue. So, again, let me not in the following restrict myself to tragic dilemmas in my definition of moral dilemma.

suggestion made by Rawls that what is appropriate is 'regorse'. (See Sinnott-Armstrong (1996), p. 55). Something which is not quite remorse, but not merely regret either.

¹³⁴ What is called perplexities *Secundum Quid* in the literature. See Sinnott-Armstrong (1988), p. 102-7.

¹³⁵ Sinnott-Armstrong, (1988), p. 52, my italics.

¹³⁶ There are extreme counterexamples, though. I break a promise to meet for dinner with some morally relevant justification. If it turns out that if I had kept the promise, both me and my dinnerdate would have been dead – say because of a bomb in the restaurant – it does seem awkward to apologise, to explain later and especially to express regret. See Foot (1995), p. 123 and Railton (1996), p. 153-54. This however does not count seriously against the point I am making here, I think.

7. Matched moral reasons and a revised definition of moral dilemmas.

One other feature of Sinnott-Armstrong's definition will prove very useful in the following, however. That is his use of the notion of a *non-overridden* moral requirement.¹³⁷ To understand this notion, we must notice that words like 'moral requirement', 'ought', 'duty', 'obligation' and 'reason' are ambiguous in everyday language.

For example, if I utter the sentence 'You ought to ϕ ', you may take me to mean that ϕ -ing is *the* right thing to do – that is, you may take me to be making an all things considered practical verdict as previously defined. Or you may take me to mean that there is *some* moral reason to ϕ , though, of course, there may be other stronger moral reasons not to ϕ . Or to take the concept of a moral requirement: Judging that x is morally required to ϕ can be taken to mean a) that x has a moral requirement *that may or may not be overridden* to ϕ . A *prima facie* (or *pro tanto*) requirement, in other words. But, it could also mean b) that x has an *overriding* moral requirement to ϕ or as some people in the debate put it: that x *must* ϕ .

Now – and here I argue in parallel with Sinnott-Armstrong – neither reading of 'moral reason' gives us what we want in our quest to define moral dilemmas. Defining dilemmas as conflicts between possibly overridden moral reasons is clearly unsatisfactory. Such conflicts, that is, may be resolvable – namely when one of the reasons overrides the other – and hence they are not guaranteed to be dilemmas. Defining moral dilemmas as conflicts between overriding moral reasons, on the other hand, is contradictory. By definition, two overriding moral reasons cannot conflict, since being an overriding moral reason means having beaten *all* other relevant reasons. As a consequence of this, we should rather want to define a moral dilemma as a conflict between *prima facie* moral reasons of a certain kind. That is, moral reasons that, on the one hand, are neither over-ridden nor outweighed by stronger moral reasons and that, on the other hand, are not themselves overriding or outweighing all other moral reasons. For short, we can say that dilemmas are conflicts between *matched moral reasons* (echoing Sinnott-Armstrong's conflicts between

¹³⁷ Sinnott-Armstrong's proposal is developed (and modified) in Brink (1996) and Railton (1996).

non-overridden moral requirements, but without his implicit restriction to tragic dilemmas).

With all this aboard, I can now attempt to give a (slightly) improved version of the definition of a moral dilemma:

Any situation is *a moral dilemma* where at the same time:

- (1)* There is a moral reason for an agent to adopt each of two alternatives
- (2)* Neither moral reason is overridden or outweighed in any morally relevant way.
- (3)* The agent cannot adopt both alternatives together.
- (4)* The agent can adopt each alternative separately.

At this point, I will make a short digression to acknowledge another potential ambiguity in the use of the concept of a moral reason in this definition. One might here read 'reason' as being about *ideal* reasons – that is, reasons defined as reasons for a fully informed and fully rational deliberator. Or one might take 'reason' here to mean what we might call an (*epistemically*) *agent-relative* reason – i.e. a reason that is relative to a particular, perhaps imperfect, informational state as it is accessed by a, perhaps imperfectly rational, agent.

Now, it would seem that in everyday discourse, it is sufficient for finding oneself in a moral dilemma that one has *agent-relative* reasons to do both of two incompatible things. It would seem, that is, that the having to act in an uncertain situation – in a situation where perhaps you know that your informational state is less than complete – is sufficient for counting the situation as a moral dilemma. The agony of facing a dilemma is sometimes simply that you have to choose while at the same time having doubts about what you ought to choose. Perhaps, you have time to dither for a while, but in the end you have to plump – to use Blackburn's¹³⁸ terminology – perhaps knowing that more deliberation and more information would have allowed you to form a rational opinion about what you have overriding reason to do. To conceive of dilemmas this way, we would have either to make clear that we read 'reason' in the definition above as agent-relative reason or to reformulate (2)* above:

- (2)*' For all that the agent has so far been able to ascertain, neither moral reason is overridden or outweighed in any morally relevant way.

¹³⁸ See Blackburn (1996), p. 127-29.

A good way to put this is perhaps to invert the following quote from Bernard Williams. Speaking of Agamemnon agonising over having had to kill his daughter to live up to his duty as commander in chief of the Greek army, he wrote:

‘..[H]e lies awake, not because of a doubt, but because of a certainty’.

‘The agonies that a man will experience after acting in full consciousness of such a [dilemmatic] situation are not to be traced to a persistent doubt that he may not have chosen the better thing; but, for instance, to a clear conviction that he has not done the better thing because there was no better thing to be done.’¹³⁹

My point has been that the agony of moral dilemmas can derive exactly from this fear that one has been forced to act prematurely – forced to act before having found the overriding reason to act in one way or the other and therefore to act while one may have overlooked things, been prejudiced and so on. The agony of a moral dilemma can be that one is not sure that one ended up doing what one had an overriding moral reason to do. This is perhaps also the reason why feeling guilt or remorse comes easy after the practical resolution of a dilemma – though these feelings, as I have briefly tried to argue, are as unjustified as blaming others for having plumped the other way in a hard case.

However, this latter point of clarification of our definition of a moral dilemma will, as far as I can see, not play any important role in answering the question to which I now turn: Can the moral cognitivist, now armed with a (relevantly) refined definition of moral dilemmas, answer the challenges of inconsistency posed by Williams and Wiggins?

8. Matched moral reasons and inconsistency

The crucial point to make in response to both challenges is that, when moral dilemmas are defined in terms of matched moral reasons, the logic of moral dilemmas will be

¹³⁹ See Williams, (1965), p. 173. Williams has, since, warmed to a more cognitivist interpretation of moral discourse. See Williams (1995b) and may therefore have changed his mind about dilemmas too.

such that they do not imply any inconsistency. They do of course imply *act-inconsistency* in that the actions the matched moral reasons jointly recommend are incompatible. But, there is no *truth-inconsistency* involved. To see this, notice that in a moral dilemma – on our Sinnott-Armstrong-inspired account – we have a situation where x believes that he *ought to* ϕ and that he *ought not to* ϕ . The point is now that this does not collapse into a formal contradiction of the form p and $\text{not-}p$ – that is, we do not get: ‘ x ought to ϕ ’ and ‘It is not the case that x ought to ϕ ’. At least not as long as the ‘ought to’ here is interpreted as ‘has a matched moral reason to’.

Let me show this by showing that the two main arguments for the inconsistency of moral dilemmas fail to go through on our revised version of Sinnott-Armstrong’s characterisation of moral dilemmas. The standard definition of moral dilemmas (reading the diamond here as physical or moral or logical ‘can’, and reading ‘A’ as ‘ x ϕ -es’ and ‘B’ as ‘ x ψ -es’) gives us:

1.	OA	Standard
2.	OB	Definition
3.	$\neg\Diamond(A\&B)$	of Moral Dilemma

This yields a ‘disagreement’ of the form: OA and $O\neg A$, if we accept the plausible principle, P3, of closure under ‘can’ (I take this to be highly uncontroversial):

4.	$(OA \& \neg\Diamond(A\&B)) \rightarrow O\neg B$	P3: Closure under ‘can’
5.	$O\neg B$	1,3,4 $\rightarrow E$
6.	$OB \& O\neg B$	2,5 $\&I$

However, given two apparently plausible principles of standard deontic logic, P1 and P2 below, any proposition of the form of (6) gives us a contradiction:

7.	$OB \rightarrow PB$	P1: ‘ought’ implies ‘permissible’
8.	$PB \rightarrow \neg O\neg B$	P2: the standard definition of permissible
9.	$OB \rightarrow \neg O\neg B$	7,8 transitivity of \rightarrow
10.	$\neg O\neg B$	6,9 $\&E$ and $\rightarrow E$
11.	$O\neg B \& \neg O\neg B$	6,10 $\&E$ and $\&I$

However, since – on our account – moral dilemmas are formulated as conflicts between *matched moral reasons*, the principles P1 and P2 will *not* give us the contradiction in line (11). The point is that no consistent interpretation of the terms of these principles make them true simultaneously. If we keep the interpretation of O

constant as ‘there being a matched moral reason to’ (as we must according to our definition of moral dilemma), there will be three cases to consider according to how we interpret P – the permissibility-operator:

If P is interpreted as: ‘there is no matched moral reason not to’, then P2 is trivially true, but P1 is false. P1 is false, since it is not the case that, if I have a matched moral reason to choose alternative B, then I do not have matched moral reason not to choose B. Moral dilemmas are a counterexample to this version of P1.

If P is interpreted as: ‘there is no overriding moral reason not to’, then P1 is true. If I have a matched moral reason to choose alternative B then it is certainly not the case that I have an overriding moral reason not to choose B. But, this time P2 becomes false. It is false that if I do not have an overriding moral requirement not to choose B then I do not have a matched moral reason not to choose B.

Finally, if P is interpreted as: ‘there is no moral reason not to’ then P1 will be false (no matter how O is interpreted, in fact) on any account that allows for moral conflict and P2 is true.

In sum, the argument from closure under ‘can’ fails to show that there is any truth-inconsistency involved in moral dilemmas understood as conflicts between matched moral reasons. But, it is also very important to notice and remember in the following that the argument *does go through when O is interpreted in terms of overriding moral reasons* and P in terms of overriding or matched moral reasons. So, if moral dilemmas had been formulated in terms of overriding moral reasons, there *would* be truth-inconsistency.

The second standard argument for the truth-inconsistency of moral dilemmas fails to affect Sinnott-Armstrong-style dilemmas for similar reasons. The argument from ‘ought’ implies ‘can’ relies on the agglomeration-principle:

1.	OA	Standard
2.	OB	Definition
3.	$\neg\Diamond(A\&B)$	of Moral Dilemma
4.	$O(A) \rightarrow \Diamond(A)$	P4 "Ought" implies "can"
5.	$O(A) \& O(B) \rightarrow O(A\&B)$	P5 Agglomeration-principle
6.	$O(A\&B)$	1,2,5 \rightarrow E
7.	$\Diamond(A\&B)$	4,6 \rightarrow E
8.	$\Diamond(A\&B) \& \neg\Diamond(A\&B)$	3,7 &I

But, the agglomeration-principle does not hold for matched moral reasons. Sophie's dilemma is a counterexample, as Sinnott-Armstrong points out. Sophie does have matched moral reason not to choose each of her children to go to the gas-chamber. But, does she therefore have a matched moral reason to choose neither? The answer is no, since the moral reason to choose neither child *is* overridden by the reason for avoiding that both get killed by the Nazi guard. Notice again that the principle, P5, *does* hold for overriding moral reasons and thus that inconsistency will be implied by conflicts between overriding moral reasons.

To sum up these considerations about Sinnott-Armstrong's proposal and truth-inconsistency, let me finally look at two sentences relating to the example of Sophie's choice: 'Sophie ought to choose her eldest child' and 'Sophie ought to choose her youngest child'. The point is that no matter how we interpret 'ought' in accordance with our revised version of Sinnott-Armstrong's vocabulary, we will get a *determinate* verdict on their truth or falsity. If 'ought' is interpreted as 'has a moral reason' or 'has a matched moral reason', then the sentences are certainly true in a dilemmatic situation. If 'ought' on the other hand is interpreted as 'has an overriding moral reason', then they are certainly false.

This is essentially the reply to the two challenges to cognitivism from moral dilemmas as I shall try to make clear in the remaining part of this chapter. To anticipate: on our account dilemmas do not imply truth-inconsistency, although they do imply what might be called moral act-inconsistency – understood as a failure on the part of moral reasoning to give unique advice, to identify *the* act to perform in all situations.

9. An answer to Williams' challenge.

Given these observations about the logic of matched moral reasons, we can easily see why Williams' argument will fail to establish that moral thinking must be inconsistent in cases of conflict and dilemma. Both when we are describing a moral conflict generally – cases where *prima facie* moral reasons conflict – and when we are describing the special case of moral dilemmas – cases where matched moral reasons conflict – we need not fear that we will come to make inconsistent judgements. It is simply not contradictory to hold that someone both has a (possibly overridden) moral

reason to ϕ and a (possibly overridden) reason to not- ϕ . Nor is it contradictory to hold that someone has a matched moral reason to ϕ and a matched moral reason to not- ϕ .

However, let us not let the cognitivist off the hook too easily. One might very well be worried about changing the definition of moral dilemmas in the way suggested here – the way that saves us from truth-inconsistency – for the following reason. One might hold that the very point of practical reasoning – understood as the pursuit of truths of the form: All things considered, x ought to ϕ in C – is to guide action. It is to help us identify an unique action in every situation – *the* right thing to do. However, if we buy the Sinnott-Armstrong-style characterisation of moral dilemmas in terms of matched moral reasons, we will lose this feature of practical reasoning. On his and our account, we have to accept that there can be more than one winner in the competition for being the right action. Our moral theory will be act-incomplete in the sense that it does not provide us with *a complete moral decision-procedure*. Let us define this notion:

A moral decision-procedure is complete if and only if, for every morally relevant action-alternative in every morally relevant situation, the procedure picks out one uniquely right action-alternative.

So, the question to ask the cognitivist is whether we ought not hold on to a logic for moral reasons that does aim at completeness in this sense. What would such a logic for moral reasons look like? Well, the following principle would disallow the existence of merely matched moral reasons:

(A) If there is a moral reason to ϕ , and if this moral reason is neither overridden nor outweighed in any morally relevant way, then there is an all things considered (or overriding) reason to ϕ .

This principle ensures that there is a constant incitement in our moral reasoning towards finding a conclusive verdict of the form: x ought all things considered to ϕ – a verdict that uniquely identifies one action as the right thing to do. The reason is simple. If the principle (A) holds true, then the two 'oughts' that I am faced with in a dilemma will be of the overriding sort. And, as we saw in the previous section, both

the arguments from dilemmas to inconsistency go through when standard deontic logic – the logic of overriding or all things considered reasons – is adopted. So, my practical thinking will be inconsistent, until I find considerations that allow me to point to a winner – to reach a conclusive practical verdict and to identify a unique action to be chosen.

Let us follow Blackburn in calling this type of logic for moral reasons a *verdictive logic*.¹⁴⁰ The question is now, how the cognitivist can argue against the adoption of this type of logic. There seems to me to be at least three good reasons for not accepting a verdictive logic for the type of moral reasons referred to in the definition of moral dilemmas.

First of all, it seems to me that the principle (A) has little intuitive appeal. In fact, it is quite hard to make sense of. One way to try make sense of it should be mentioned. If one adopted an *absolutist* deontologist moral theory according to which there are moral principles that must be followed on every occasion with absolutely no exception – absolute moral duties that *no* stronger moral duties could *ever* override – then a moral dilemma could be conceived of as a conflict between overriding moral reasons and something like the principle (A) would hold true. Thus, we can easily imagine cases where no matter what one does, one would violate what a theorist of this type considers to be an absolute duty – say a case where both available options involve killing.

However, this type of absolutist moral theory in my opinion in itself makes little sense. Why should we accept the existence moral reasons of such supreme weight? Why not allow for the possibility of ranking reasons? It seems to me that if we accept the very plausible principle that any moral theory should somehow be based on a theory of good or value – even the deontologist should accept this in the sense that his moral principles are designed to guide us to *respecting* moral values – then we should allow for the possibility of ranking, at least in extreme cases.¹⁴¹ So, there is a problem of making sense of this kind of extreme absolutism and in a wider context of making sense of any theory that asks us to adopt the principle (A).

¹⁴⁰ See Blackburn (1996), p. 134-37. Blackburn here argues against Sinnott-Armstrong's attempt to ensure truth-consistency in moral reasoning. Perhaps he, as a non-cognitivist, is less worried about inconsistency in moral reasoning.

¹⁴¹ This point is well argued for in Griffin (1996), e.g. p. 110-111.

Secondly, it should be noticed that, although moral theory involves act-incompleteness in the sense defined above, there is a weaker sense of completeness according to which our theory is complete. Thus, on our characterisation of moral dilemmas, there will for any action-alternative, A, be a determinate answer to a question of the form: Is it morally permissible (or impermissible) to choose A? It is just that in some situations, we will answer yes to both the question about A and about not-A. Hence, there will be no moral judgement the truth-value of which would be somehow indeterminate. And, I hold, this is the fact that is relevant to making a decision about the objectivity of moral discourse.

It seems to me that the demand that moral reasoning should give us a complete moral decision-procedure – as opposed to a weakly complete decision-procedure – is just as unreasonable as the demand that one should always come up with a unique number when answering arithmetical questions. Surely the fact that 'What is the square-root of 4?' has the answer '2 and minus 2' tells us nothing about the objectivity or truth-aptness of arithmetical discourse. Equally, the fact that both ϕ -ing and not ϕ -ing may be morally permissible actions in the case of a moral dilemma tells us nothing about the objectivity or truth-aptness of moral discourse. I shall return to this point in my answer to Wiggins' challenge.

Thirdly and finally, it would seem that to insist on there always being an overriding reason for one alternative in the case of a dilemma – to insist on a verdictive logic of moral reasons – rests on the heavy metaphysical assumption mentioned in section 3 of this chapter. The idea, that the universe is somehow designed such that ideal moral agents could avoid moral conflict and dilemma and identify one uniquely right action to perform in every situation. Given that we have no reason to accept this assumption, why not reflect this fact in our moral reasoning by adopting the logic of matched moral reasons advocated here?

In conclusion, it does seem that we have found a way of representing the real phenomena of moral conflict and dilemma such that it does not compromise the cognitivist's claims. We have answered Williams' challenge on behalf of the cognitivist.

10. *An answer to Wiggins' challenge: An unsuccessful attempt.*

Recall now Wiggins' challenge:

1. If there are absolutely un-settleable moral disputes and if moral truth is epistemically constrained then moral cognitivism is inconsistent.
2. Moral dilemmas are examples of absolutely un-settleable disputes.
3. There are moral dilemmas.
4. Moral truth is epistemically constrained.

∴ Moral cognitivism is inconsistent (hence false).

Given the way we reasoned in the previous chapters, it would seem that we should immediately grant Wiggins premises 1 and 4. Now, we have already encountered reasons for doubting that moral dilemmas present us with absolute un-settleability in the sense relevant to moral objectivity, but let us for now grant Wiggins his premise 2 as well. When we have made these assumptions, the only possible way for the cognitivist to meet Wiggins' challenge is to deprive him of his third premise – to claim that we do not, at least not *yet*, have a warrant for the strong claim that there are moral dilemmas. This is a version of the strategy (successfully) adopted in the previous chapter. Does it work for dilemmas too?

The cognitivist's strategy would seem to have to involve justifying the claim that the third premise seems never to be fully justified – that we can remain agnostic about the existence of moral dilemmas. What we do have – the cognitivist would have to argue – is evidence for *de facto* un-settleability of certain inner moral conflicts. But, to justify the claim that the conflict is *absolutely* un-settleable, we must somehow convince ourselves that no way of turning the evidence and the moral principles involved in our heads – including no way of re-conceptualising the whole structure of moral reasoning – will bring about a resolution. But, the cognitivist would ask, When is this ever likely to happen? How could we ever come to justifiably believe that we had exhausted all possible ways of finding some way of ranking the reasons involved, viewing the evidence in a novel way or... such that we would be able to resolve the issue?

If what the cognitivist argues here holds water, it would seem that adopting an agnostic attitude towards the question of the existence of moral dilemmas – and hence to the claim about moral completeness – is reasonable. And then, since: (1) EC, (2)

agnosticism about (the *a priori* status of) moral completeness, and (3) agnostic about (the *a priori* status of) bivalence is a consistent triad, there is a consistent version of cognitivism that can deal with the phenomenon of un-settleability, including here the phenomenon of irresolvable conflict between moral reasons for one agent at one time.

But, does the cognitivist's agnostic answer hold water in the special case of moral dilemmas? Agnosticism seemed to be an attractive answer when we considered radical moral disagreements. But, it would seem that moral dilemmas pose a special threat. That is, that we *can* conceive of definite cases of genuine moral dilemmas. Let us try to assess the claim that we can remain agnostic about the existence of moral dilemmas by briefly considering the two standard ways of arguing for their possibility and actuality. The first argues from the plurality of incommensurable values and principles to the incomparability of some moral reasons. The second argues from the possibility of absolutely symmetric moral reasons deriving from the same moral principle or value.

The first strategy does seem to allow for the adoption of an agnostic attitude. Consider the example of someone torn between a professional promise and a special obligation not to hurt his daughter. The agent here cannot fulfil his promise without missing a special occasion at home with his daughter. Now, the argument of the defender of moral dilemmas need not here be that there are *no* cases where moral reason of the types involved are comparable. A moral reason to keep a very important promise may clearly override an obligation to avoid a slight discomfort for one's daughter – and thus the requirements encountered in this case are comparable. But, the defender of the possibility of moral dilemmas may plausibly advance only the limited claim that *some* particular moral reasons of one type do not compare with *some* particular moral reasons of another type.

But, any claim about the impossibility of ranking moral reasons – whether it is the ranking of types or of particular instances of moral reasons – are surely defeasible claims. It *does* make sense to raise the prospect that future moral theory might include some way of ranking the moral reasons such that a resolution of the dilemma would come about. Surely, it is an everyday occurrence that one comes to realise that one must change one's value-conceptions such that values that one did not previously rank, one does now rank.

However, the second strategy for arguing for the existence of moral dilemmas – the case of absolutely symmetric cases – does appear to cause a problem for the agnostic cognitivist. Think of Sophie's Choice. If (what is surely conceivable and possible though perhaps never actual) there is absolutely no morally relevant way to distinguish between the two children (they are twins, they have exactly the same life expectancy etc.) then it seems not to the point to claim that maybe some future philosophy will give us an answer to whom Sophie all things considered ought to choose. It is in fact conceivable that there are dilemmatic cases that absolutely no moral consideration could ever decide. And of course it does not help to argue that it is *merely possible* that dilemmas of the symmetric sort can occur. Admitting to the possibility of inconsistency is just as bad as its actuality.

In other words, it would seem that – due to the possibility of symmetric moral dilemmas – we also have to grant Wiggins his third premise. But then – at least in so far as we continue to work under the assumption of the truth of his second premise – we must also accept his anti-cognitivist conclusion. It seems that the acceptance of the argument from symmetry to the actuality or possibility of moral dilemmas gives us a strengthened version of Wiggins' challenge to contend with. What can be done?

11. Two answers to Wiggins' strengthened challenge.

One answer here would obviously be to give up the agnostic attitude towards moral completeness and admit that there are absolutely un-settleable moral questions and, hence, that moral reasoning is definitely incomplete. This radical solution, however, ultimately amounts to the elimination of the minimalist's position from the debate. The minimalist would then hold true three jointly inconsistent claims about moral discourse: (1) EC, (2) Incompleteness and (3) the double-negation of the Law of Excluded Middle – as Wiggins' argument shows.

Now, one reaction to the symmetric cases, seemingly available to the cognitivist, is to *go agnostic about EC*. But, this does not get rid of the inconsistency.

1	(1)	$\Diamond Kp \leftrightarrow p$	EC
2	(2)	$\neg(\Diamond Kp \vee \Diamond K\neg p)$	Dilemmas (as Incompleteness)
1,2	(3)	$\neg(p \vee \neg p)$	1,2, substitution ($p/\Diamond Kp, \neg p/\Diamond K\neg p$)
	(4)	$\neg\neg(p \vee \neg p)$	Intuitionistic (& classical) theorem
1,2	(5)	\perp	3,4 $\neg E$
2	(6)	$\neg(\Diamond Kp \leftrightarrow p)$	5, $\neg I$

Here it has been shown that Dilemmas (understood as involving incompleteness) can be used to prove the falsity of (EC). But, it cannot then be possible to remain agnostic about (EC) – since one can obviously not be agnostic about what one can prove the negation (or affirmation) of. Accordingly, moral minimalism – understood as the claim that moral discourse is truth-apt, yet fails to satisfy the Cognitive Command constraint – coupled with agnosticism about the failure of EC, is not a consistent position in the face of the actual (or possible) existence of moral dilemmas. So, we are back with the choice between no truth in ethics (non-cognitivism) or strongly objective truth (cognitivism with Cognitive Command) in ethics. Moral minimalism disappears from the map of the debate.

But, I want to argue, this first answer to Wiggins' strengthened challenge is an overreaction anyway. The dialectic of the previous section and the beginning of this section rested on the assumption that we have to grant Wiggins his second assumption. That moral dilemmas are properly seen as cases involving un-settleability in the relevant sense. But, this assumption has already, at least implicitly, been rejected in the foregoing. Let me now answer Wiggins' strengthened challenge by making it explicit.

The point to make is of course that the sense in which moral dilemmas are un-settleable does not entail the type of incompleteness that causes problems for the cognitivist. As mentioned earlier, the existence of moral dilemmas as we characterise them implies, on the one hand, that moral reasoning gives us only an *incomplete decision-procedure*. But, why should we expect more? On the other hand, the existence of moral dilemmas *is* compatible with the *weak completeness* of moral reasoning. And it is the latter, I argue, that is important here. To appreciate this point, let us finally look more carefully at Wiggins' argument. As I interpreted him, he thought that – in the face of un-decidable moral dilemmas – we would have to assert the following:

'The denial of '(A would be right) or not (A would be right))'.¹⁴²

A similar – mistaken according to me – claim is made by Russ Schafer-Landau. He argues in relation to the Sophie's Choice example:

'Though it may have been determinately true that [Sophie] was obligated to sacrifice one of her children, *it may have been neither true nor false that she was obligated to choose her eldest child*' (my italics).¹⁴³

The question to ask both Wiggins and Schafer-Landau is, what could 'is right' and 'having an obligation to' mean in these quotes, if the quotes are to be true? I will now argue that, as a matter of fact, there is no moral p such that p 's being the focus of a moral dilemma leads to a denial of LEM. A moral dilemma – on our characterisation – will lead us to make the following 5 verdicts:

- (1) It is true that: There is an overriding reason to (ϕ or ψ)
- (2) It is true that: There is a matched reason to ϕ
- (3) It is true that: There is a matched reason to ψ (hence not to ϕ)
- (4) It is false that: There is an overriding reason to ϕ
- (5) It is false that: There is an overriding reason to ψ (hence not to ϕ)

This is just the upshot of the claim that moral reasoning, when adopting our non-verdictive logic for moral reasons, is weakly complete.

Now, if 'is right' and 'being obliged to' here means something similar to what 'having a matched reason to' means, then it is *true* that A would be right (to use Wiggins' symbolism) and so Wiggins' suggestion that both 'A would be right' and 'Not(A would be right)' are false is not vindicated. If 'is right' and 'being obliged to', on the other hand, is thought to mean something like 'having an overriding reason to', then 'not(A would be right)' is true and so again Wiggins' suggestion that both disjuncts are false is not vindicated. Notice also that there is nothing contradictory in claiming (4) and (5) together, although, of course, they would have been mutually

¹⁴² Wiggins (1998), p.131.

¹⁴³ See Russ Schafer-Landau (1994) and (1995). He goes on to argue that this is *not* a problem for moral objectivism or realism.

inconsistent if they had been: 'It is true that there is an overriding reason to ϕ ' and 'It is true that there is an overriding reason to ψ ' as we saw above. To say that there is *no* overriding reason to ϕ is just to say that it is permissible not to ϕ . And there is nothing contradictory in saying that both ϕ -ing and not ϕ -ing are equally morally permissible.

Hence, moral dilemmas do not give us a reason to believe in moral incompleteness. Moral reasoning is (weakly) complete in the sense that there are definite verdicts of the form (1) – (5) above to be made when faced with a dilemma. Of course, we have to admit that moral reasoning does not always identify a unique right action. Moral reasoning alone does not always suffice. But, what we called decision-procedure incompleteness is a different matter altogether. Moral minimalism is still on the map.

Chapter Five

Minimalism and The Location Problem.

Moral minimalists accept that there are moral facts. That much follows trivially on accepting that moral discourse is truth-apt and that not all moral judgements are false. That is, it follows on *not* embracing a non-cognitivist or an error-theoretic account of morals. But, this would appear to commit the minimalist to some sort of special moral ontology. The question then becomes: How do moral facts – moral entities and properties – fit into our overall ontology? Here the minimalist appears to be faced with two options. His first option is to claim that moral facts are *sui generis*. And the second is to hold that moral facts are reducible to some other realm of facts. And the candidate for being the reductive class is, of course, most often some variety of natural facts.¹⁴⁴

Now, as many in the current debate have seen the first, *sui generis*, option as wholly untenable, the moral cognitivist – including the minimalist – is regarded by many as being compelled to choose the second option. That is, many in the debate have held that treating morality as *sui generis* compels one to adopt a very strong form of moral realism – call it 'Platonism' or in John McDowell's more evocative terms 'rampant platonism'.¹⁴⁵ According to Platonism, moral judgements are made true by an autonomous reality of moral facts – a reality that is autonomous in the sense that it bears no essential relation to human beings and their needs and practices. Moral facts are, according to this view, robust – as robust as are, for example, natural facts about primary qualities – yet they belong to some *supernatural* realm. Now, since the *sui generis* option has been seen as linked to 'rampant platonism' in this way – and since 'rampant platonism' has been generally discredited, perhaps most famously by Mackie's argument from queerness against moral realism – the

¹⁴⁴ Natural facts are for the purposes of the rest of this chapter to be understood as facts that can be fully described in naturalistic language (which of course leaves open the possibility that they can be described in the language of morals too). Naturalistic language, in turn, is the kind of language that can be used to give causal explanations of the kind that feature in the empirical sciences – including sociology and psychology, (i.e. language that contains predicates that can figure in the formulations of empirical laws).

¹⁴⁵ See McDowell (1994), p. 77.

naturalistic option has been the most popular among those of a moral cognitivist persuasion in the modern debate.

If the minimalist is persuaded by this account of the dialectic – the account that equates the *sui generis* option with Platonism – and, accordingly, goes for the naturalistic option, he will see the problem of moral ontology as Frank Jackson sees it – as what he calls *the location problem for ethics*.¹⁴⁶ To use his words, we then have to 'offer an account of how the ethical gets a place in the descriptive [i.e. naturalistic] picture of what our world is like'.¹⁴⁷ This choice will inevitably lead to some sort of a reductivist¹⁴⁸ account of moral discourse – to what McDowell has called 'bald naturalism'. There have been many attempts to provide some such localisation in recent years – both by analytical reductivists and by philosophers inspired by Kripke's and Putnam's postulation of the possibility of a posteriori necessary identities. I shall later – in chapter 6 – return to a discussion of some of these attempts.

However, in the following I will argue that the minimalist should not allow himself to be persuaded by this account of the dialectic – the account that sets up a false dichotomy between 'rampant platonism' and 'bald naturalism'. Treating moral discourse as *sui generis* and as apt for truth does not necessarily lead to a realistic – Platonist – account of morals exactly because minimalism exemplifies a position that allows for our notion of truth to be less robust than that of the realist – with its involvement with more robust notions of correspondence and fact. The minimalist's idea is exactly that understanding how a notion of truth can govern our practice of moral argument does not have to involve the postulation of an independently existing moral realm – a realm that we as moral practitioners are trying in some substantial sense to reflect in moral theorising. In other words, the suggestion I will want to develop in the following is that the minimalist can choose the *sui generis* option, yet avoid falling into the Platonist position traditionally associated with this option. I shall

¹⁴⁶ Jackson (1998), p. 113. Though he does not want to use the word 'natural', but prefers descriptive.

¹⁴⁷ Jackson (1998), p. 113

¹⁴⁸ I use 'reductivist' here in a broad sense that include theories that are in a narrower sense non-reductive. I am thinking here of naturalists (non-reductive in the narrow sense) that deny the possibility of *analytical* reductions, but affirm synthetic identities between moral and non-moral properties. Thus, when for example the Cornell realists claim that they treat morals as *sui generis*, they do so because they deny the possibility of analytical reduction, not a posteriori identities (See Railton (1993b), p. 317 and Darwall, Gibbard, Railton (1997), p. 26-27). Hence, they use 'sui generis' in a broader sense than I do, since I include Cornell-realists among the reductivists (since they are substantive naturalists).

call this proposed way out *minimalist platonism* (or *platonism* with a small ‘p’ or simply *minimalism*).

This account of the dialectic poses a set of tasks for the minimalist. First of all, he must make clear that his account of morals does not slide into Platonism. Minimalism ought to make naturalistic sense in a broad sense of this word – there should be no metaphysically extravagant postulation of a mysterious moral realm. A stable position in between ‘bald naturalism’ and Platonism must be found. Secondly, this minimalist platonism must provide us with the resources to take on at least three further tasks – namely exactly the tasks that Platonism does not fare well with respect to: a) giving an account of moral epistemology, b) giving an account of the normativity or directedness of moral judgement, c) giving an account of the tendency of moral beliefs to motivate.

In this chapter, I attempt to perform the first two of these tasks on behalf of the minimalist. I have, of course, already addressed the question of the tendency of moral beliefs to motivate in the second chapter. The defence of minimalist platonism is structured as follows. I attempt to show that the minimalist can opt for platonism without turning to Platonism – avoiding queer facts (section 1). I argue that the platonist can meet the epistemological challenge that defeated the Platonist without compromising the claim that moral facts are mind-independent in the relevant sense (sections 2 and 3). And, finally, I argue that the platonist has no problem accounting for the normativity of moral judgement (section 4). In the next chapter, I then investigate further the dialectic between moral minimalism and ‘bald naturalism’.

In the following two sections, I say more about another form of naturalism – an explanatory or methodological naturalism attempting to give naturalistic explanations of the existence and character of truth-governed moral discourse. Crucially, this form of naturalism *is* compatible with moral minimalism (section 5 and 6). I then turn to the issue of moral relativism and the question of whether minimalism is to be seen as a form of relativism. I define different kinds of relativism (section 7). And I argue that a strong form of relativism – that I shall follow Gibbard in calling ‘parochialism’ – is untenable and incompatible with a methodologically naturalist understanding of the function of moral discourse (section 8). And I argue that this gives moral minimalism a dialectical advantage over the parochialist (section 9).

1. Platonism, platonism and queer facts.

There is no doubt that 'rampant platonism' *does* fall prey to Mackie's famous argument from queerness against moral realism.¹⁴⁹ The 'rampant platonist' believes in the existence of a supernatural moral realm – features or entities that are not part of the natural world, yet are as real as natural features and entities. But, it is easy to point to the seemingly insurmountable metaphysical and epistemological problems with any such view. Firstly, moral states of affairs are such that, when we believe they obtain, they give us reasons to act and they tend to motivate us accordingly. Following the Platonist's account of morals, this means that there are entities or features in the world that are somehow intrinsically reason- and motivation-giving when known – queer entities or features quite unlike any other entities or features in the world. Secondly, if there are such supernatural entities, then the means by which we come to know them cannot be the same as the means by which we acquaint ourselves with natural facts – sensory perception, that is. In other words, the Platonist must postulate a special faculty to detect these queer facts – along the lines of the proposals of the moral intuitionist. And this ability would equally be queer – quite unlike any other of our epistemic faculties.

One does not have to be blinded by scientism and empiricism to share Mackie's unease with the idea of an independent moral realm – wholly outside nature, *yet as real in every sense as nature* – and of special moral faculties. The Platonist's dualism puts him at odds with the modern scientific approach to nature in ways that are just unacceptable to anyone who is suitably impressed by the fruitfulness of adopting the approach of modern science. 'Rampant platonism' *is* unacceptably queer according to most philosopher's ontological taste (including mine). But, quite apart from this argument – which is perhaps just a reflection of ontological prejudice – there is Ockham's razor. If there is an adequate account of moral practice that has lighter ontological commitments, then we ought to choose that account over the Platonist's. And the minimalist's claim is that such an account is in the offing.

But, the potentially troublesome questions to the minimalist are: Is Platonism not just the consequence of accepting the truth-aptness and the irreducible nature of

¹⁴⁹ Mackie (1977), p. 38-42.

moral discourse? Is there really such a thing as a minimalist platonist position distinguishable from Platonism? And, if so, can *it* avoid Mackie's queerness-objections. In this section and the next, I will attempt to answer these question.

What is distinctive of the minimalist's form of platonism is his minimalistic notions of truth and truth-aptness. What is required for a concept's being a concept of truth according to the minimalist is merely that it satisfies a set of platitudes. And what it takes to be truth-apt is just that appropriate syntax and a degree of discipline is displayed. And as we have seen, moral discourse *is* disciplined – affectively grounded moral beliefs are defeasible in various ways. In other words, the practice is already 'up and running' in the right way and this is sufficient for its being disciplined and, hence, minimally truth-apt. Crucially, this minimalistic conception of moral truth-aptness does not involve any essential play with the idea of representation of, or correspondence to, some radically external reality (other than in the merely trivial sense in which any true statement may be said to 'fit the facts' or 'tell things as they are' as a matter of idiomatic convention). The discipline in question need not be the kind of discipline that derives from an attempt to correctly represent something independently constituted – where getting it right means representing correctly in a *substantial* sense. There need be no ontology of the relevant kind independently of – conceptually prior to – the possibility of setting up a corresponding practice. Indeed, the central minimalist idea is that things may be exactly the other way round. Ontology may be a product of discipline. An 'up and running' practice may *project* entities and properties of the relevant sort onto our world.¹⁵⁰

So, the moral minimalist's position is that something like this is true of moral discourse: Moral reality is in a sense constituted by human moral practice based on (educated) affective responses of its practitioners. But, where does that leave the minimalist with respect to fitting moral ontology into our overall ontology. The answer must here be that the very notion of *ontology* can be taken in two ways. There is a broad conception of ontology which includes everything that is referred to in true judgements formulated in truth-apt discourse. So, in so far as we are working with

¹⁵⁰ Using the metaphor of projection here is potentially confusing, since this has unwanted connotations in modern metaethical debate. That is, it is usually linked to Blackburn's (and others') expressivist account of ethics. However, the metaphor does seem appropriate to use in the context of describing minimalism as long as it is kept in mind that what is being projected according to the minimalist are not non-cognitive attitudes, but (minimal) beliefs.

this permissive conception of ontology, we will allow that entities and properties as diverse as bikes, atoms, numbers, weight, redness and moral rightness all exist. Natural entities and features can in this sense co-exist with abstract entities and moral properties with no need for providing a reduction of the latter to the former. There is no location-problem in Jackson's sense, since *being* in the sense relevant here is just a question of *being referred to in true statements made in properly disciplined (indicative) discourse*.

But, even if this permissive notion is the minimalist's preferred notion of ontology, he must also admit that he cannot prevent others (even himself) from working with a more demanding conception of ontology and existence – say, ontology and existence – according to which only entities and properties referred to in discourse that is *apt for truth in a more substantial sense* exist or are part of our ontology.¹⁵¹ A plausible suggestion here would be that only discourses that satisfy the Wide Cosmological Role constraint are truth-apt in this more substantial sense. This admission is parallel to the admission that minimalists can accept a distinction between minimal and robust beliefs as discussed in chapter 2. The minimalist can, of course, equally accept distinctions between facts and facts or states of affairs and states of affairs. Perhaps this will mean that only entities and properties referred to in naturalistic discourse – scientific discourse – will enter our list of things that exist. Perhaps, that is, we will adopt a naturalistic ontology. When working with this conception of ontology, we may want to say things like: ‘Moral values do not *really exist* out there in nature’. But then, again, there is no location-problem. In this more demanding sense, properties like being right and being valuable do not exist and there is, hence, no problem of how they relate to natural properties.

If this line of argument holds water, then it would seem that the minimalist platonist *does* have an answer to the first part of Mackie's queerness-objection. Mackie seems to be presupposing in setting up his argument that a moral platonist must be a Platonist – i.e. be somebody who claims that moral properties are included in our ontology in the restricted sense of this word – ontology. If this assumption were true, the platonist would have to treat moral properties as he treats robust entities from our ontology. And in that context – things like atoms, aeroplanes and

asteriods – moral properties *do* appear queer. But, of course, the minimalist platonist only claims that moral properties are part of our ontology in the more permissive sense – and, in that context, there is nothing queer about them.

2. *Minimalist platonism and the epistemological challenge*

So, the platonist does not have to contend with queer facts. But, there was a second – epistemological – component to Mackie's objection. The minimalist was challenged to explain how we become acquainted with moral states of affairs. I shall now argue that the minimalist *can* answer this challenge. Let me first formulate the challenge that causes problems for the Platonist in some more detail, before showing why it does not affect the platonist.

The Platonist believes that moral facts are radically independent of human beings – that is, wholly external to human practices of moralising, but able to be represented by them. But, if the moral facts are in this way *radically external* to the subject, then, of course, we may legitimately ask: How do we come to know about this reality – how do we bridge the gap between us as knowers and the moral realm? Now, the Platonist can answer this question very easily simply by adding to his extravagant metaphysics. He can claim that we have a special *sui generis* faculty – a moral intuition – that allows us access to this acausal supernatural realm. But, this answer is available only at the cost of retaining and adding to his already extravagant metaphysics. In essence, he must claim that we humans are not entirely natural creatures. Mackie's epistemological objection to Platonism is simply – once again – that we ought not to pay the price of such extravagance.

And, of course, metaphysical extravagance is wholly unjustified – by Ockham's razor – if avoidable. The purpose of this section is to argue that the minimalist *can* avoid metaphysical extravagance. I want to defend a minimalist account of moral knowledge by arguing that knowledge of moral principles is a species of *a priori* knowledge and I want to suggest that there is no problem with simply holding on to an non-causal conception of moral knowledge. That is, the claim will be that knowledge about moral principles is warranted, true (i.e. superassertible) belief about

¹⁵¹ A plausible suggestion here would be that only discourses that satisfy the Wide Cosmological

moral principles – where warrant may, of course, be an undefeated entitlement. Also, it is claimed that knowledge of moral principles is not plausibly subjected to a strong causalist or reliabilist constraint – for the reason that moral beliefs do not aim at correspondence with substantial facts, i.e. facts with a wide cosmological role. Finally, in the next section, I attempt to meet a possible further challenge to the minimalist posed by Hartry Field. He notes first that:

‘We should view with suspicion any claim to know facts about a certain domain if we believe it impossible to explain the reliability of our beliefs about that domain.’¹⁵²

He then alleges that any *non-causal* explanation of this reliability can only be adopted on pain of giving up on the mind-independence of the given subject-matter. So, a final question to be addressed in the following is whether the minimalist can reconcile his account of the reliability of minimalist moral beliefs with a *plausible* claim about the *mind-independence* of moral states of affairs.

Before going on with this agenda, I shall argue that moral knowledge about principles should be seen as a species of a priori knowledge. We may say that a fact is knowable a priori just in case it can be known merely through introspection and/or reflection on the concepts involved – with no reliance on sensory perception whatsoever.¹⁵³ It is important to remember here the distinction between knowing particular – occasion-specific – moral truths and knowing universal moral truths – i.e. moral principles. The claim shall only be that a universal moral truth, *p*, – for instance that lying is wrong – can be known through reasoning and introspection alone.

Now, why is moral knowledge of principles a priori? If we accept the account of moral epistemology outlined in chapter one, then it should be clear that there is no need for the senses in acquiring a warrant for a belief about a moral principle. Moral beliefs are affectively grounded. So the mere fact that we spontaneously come to believe that some moral principle is true – something that we can determine introspectively and, hence, with no sensory involvement – gives us a *prima facie*

Role constraint are truth-apt in this more substantial sense.

¹⁵² Field (1989), p. 232-33. His argument was, of course, originally directed at mathematical knowledge.

¹⁵³ With the following qualification, though. Human beings cannot acquire concepts without interacting – through the senses – with a speech-community and the external world. So we must add

entitlement to have this belief. Next, notice also that realising that an entitlement like this is *defeated* is an a priori matter. According to the coherentist account of reasoning about moral principles given, our entitlement to believe that *p* (where *p* is some moral principle) is defeated when we realise that *p* is not part of the set of principles that jointly coheres with our considered and revised moral judgements about particular cases (which could, of course, all be hypothetical cases) and our even more general background philosophical principles – along the lines of reaching a Rawlsian wide equilibrium.¹⁵⁴ Crucially, we can perform these reflections without reliance on the senses. By contrast, coming to know a particular moral truth – e.g. that *this* act of lying is wrong – cannot be a wholly a priori affair, since we can only come to know *a posteriori* that some particular act is an act of lying.

So, we see how a warrant for a belief about moral principles can be attained a priori. But, how – on the minimalist account – does a warranted belief turn into *knowledge*? We must here remember that moral truth is superassertibility. Hence, if our warrant for a belief as a matter of fact is *not defeated and could not have been defeated even if we had used our actual potential for investigation to its ideal limit* – if the warrant could survive the best possible attempt in this world at defeating it – then what we believe is superassertible. That is, *true* according to the moral minimalist. And then, of course, we can be said to *know* whatever it is that we are in this sense durably warranted in believing. On the face of it, therefore, there seems to be no problem for the minimalist of making room for the idea of moral knowledge – and therefore no problem of distinguishing platonism from Platonism and answering Mackie's epistemological queerness-challenge. Thus, moral knowledge, according to the minimalist, is simply superassertible, warranted belief. (Where warrant need not be justification, but might be an undefeated affectively grounded entitlement).

A final point to make is that there is no reason to add the requirement that moral knowledge must satisfy a *strong* causal (reliabilist) constraint. Strong causal

that a priori knowledge is knowledge based on conceptual reflection alone, *given a set of already acquired concepts*.

¹⁵⁴ A wide reflective equilibrium is a coherent set of beliefs consisting of 1) considered moral judgements about particular situations, 2) a set of universal moral principles and 3) a set of more general background theories. It is reached by going back and forth between an initial set of beliefs of the three types – revising and adjusting them slightly – so that finally coherence is reached. Rawls' own account of the idea of reaching a wide reflective equilibrium is to be found in Rawls (1972), p. 20f.

theories of knowledge are theories imposing the following necessary condition on knowing: S knows that *p* only if S's belief that *p* is causally related (in the right way) to the fact that *p*.¹⁵⁵ Now, given the minimalist's framework for the debate, we should add that the notion of fact employed here should be interpreted as *substantial fact* – i.e. a fact with a wide cosmological role – in order to capture the spirit of the causalist's proposal.

It seems plausible to suggest that a strong causal constraint like this is in fact inappropriate for all kinds of a priori knowledge.¹⁵⁶ In particular, if knowledge of moral principles is knowledge that does not depend on sensory perception, then someone could come to have warranted, true belief as he reflected on moral principles even while placed in a tank that isolates him from all sensory input. Now, a sensory isolation tank is for our purposes the same as an environment where no causal input from the outside is let in. But, then it just seems odd to suppose that knowledge of moral principles must involve causal interaction with – robustly mind-independent – facts about moral principles. On the minimalist picture, moral beliefs are not candidates for corresponding to such facts. Therefore, the kind of warrant we may have for our moral beliefs cannot be constituted by a causal – or reliable – connection with such facts.

¹⁵⁵ Weak causal constraints on knowledge, by contrast, merely demand that to know *p*, the subject's belief that *p* must be caused by the subject's reasons or evidence for *p*.

¹⁵⁶ Strong causal theories are, of course, problematic not just in the case of a priori knowledge. Unless modified, it will disqualify perfectly good cases of non-moral, a posteriori knowledge – such as knowledge of the future, knowledge of general empirical truths and truths known by inference. For example, I know that all humans are mortal. But, it seems wrong to suggest that this belief is causally related to that general empirical truth. All that my belief is causally related to are individual instances of people dying. So, the strong causal account of knowledge fails to capture general empirical knowledge. See Hale (1987), p. 92-97 for the suggestion that no modifications offered so far on behalf of the causalist will manage to save all these kinds of knowledge.

But, quite apart from this worry, there is the problem that a causal constraint is not an answer to all Gettier scenarios – in particular cases that depend on false background assumptions. Think of the following example: I know that most students at the university that are native English-speakers and from the Southern Hemisphere are from New Zealand. Say that I saw some statistics in the university yearbook. Say also that I am not able to distinguish New Zealand accents from Australian accents, but that I can tell that an accent is from 'down under'. Now, I hear a somebody with a 'down under' accent on the campus and correctly, as it happens, judge that this person is a New Zealander. However, atypically, this week the ratio of Australians to New Zealanders has changed dramatically due to a large number of Australians attending a conference at the university – the fact that I happened to meet a New Zealander was pure luck. So, on the one hand, I can't really claim to know that the person in question was a New Zealander. However, on the other hand, my belief was both justified and causally related to the fact the the person was a New Zealander. That is, I came to believe that he was a New Zealander partly though interacting causally with him – i.e. hearing his accent and concluding that it is from 'down under'. See also Hale (1987), p. 87-88.

The conclusion to this short investigation is then that moral knowledge is simply true (i.e. superassertible), warranted belief. And it would seem that some sort of non-causal account of moral knowledge of principles must be defensible in virtue of the fact that knowledge of moral principles can be arrived at through conceptual reflection and introspection alone.

3. *Minimalism and counterfactual mind-dependence.*

However, it may now be argued – as it has been argued by Hartry Field – that there is a further requirement on knowledge that does not exempt acausal, a priori, moral knowledge. That is, that the believer in moral knowledge still owes an explanation of how it is that we are *reliable* in forming moral beliefs¹⁵⁷ – assuming that the minimalist does not want to go for moral scepticism.

So to pre-empt a possible line of attack against the minimalist's account of moral knowledge, we must see how the minimalist explains that moral belief-formation is on the whole reliable. The minimalist, of course, sees truth as superassertibility – as (in the already specified sense) durable warranted assertibility. But exactly for that reason accounting for the global reliability of moral belief-formation is straight-forward, it seems to me. The way moral beliefs are typically produced make them reliable indicators of moral truth because the process by which they are produced – the provision of a warrant in the shape of an undefeated entitlement – gives us exactly the warrant that (if it survives all future actually possible additions to our informational state, all future actually possible reflection on the current picture) goes on to make it true. In other words, warrant and truth are internally linked on the minimalist account and there is, therefore, a very simple explanation of why our moral belief-forming processes – i.e. warrant-forming processes – are reliable. The only thing that needs to be established is that moral subjects with the appropriate moral standards *most of the time* reason correctly (and with respect to occasion-specific

¹⁵⁷ I am inspired in the following discussion by Divers and Miller (1999). They discuss Hartry Field's attempt to strengthen the epistemological challenge to the *mathematical* platonist – I attempt to extend Field's reliabilist challenge to the moral case.

moral belief, of course, also that they have mostly true non-moral beliefs). So, it is certainly not 'impossible to explain' the reliability of moral beliefs.¹⁵⁸

But, according to Field, the minimalist is not out of the woods yet. As already mentioned, he alleges that any acausal account of the reliability of belief-forming processes – like the one just given in the moral case – must compromise the claim to *mind-independent* truth. The minimalist answer to Field must here be that there is both a sense in which moral facts are mind-dependent and a sense in which they are not. Clearly, the platonist's story – involving as it did the claim that moral facts are in a sense partly the result of the affective natures of actual human beings – does in a sense include the admission that moral facts depend on human beings. This dependence on disciplined, affectively grounded, human discourse is exactly what distinguishes platonism from Platonism.

However, there might appear to be a second sense of 'mind-dependence of moral facts' that is not compatible with platonism. Moral facts could be thought to depend on human minds in a counterfactual sense. That is, in the sense that had human beings been different – in the way they practised moral discourse – then moral truths would have been different. Or, indeed, in the sense that had there been no human beings or were the human race to become extinct, then there would be no moral truths at all.

This *counterfactual* mind-dependence, it seems, *is* incompatible with minimalist platonism. To see why, consider first of all that it is a platitude that truth is timeless – once a proposition is true it is always true and always were true. If superassertibility really is a truth-predicate, it must therefore have this property. But, consider also the following scenario. Let us say that the minds of the human subjects at time t_1 are such that p is true at that time. Imagine also that the human mind changes (or becomes obsolete) such that at t_2 it is no longer true that p – as the idea of a counterfactual mind-dependence allows for. But, then p turned out to have been

¹⁵⁸ However, might we not still imagine a community of unlucky moralisers that kept coming up with warrants in accordance with their standards that were ultimately defeated in later states of information? Of course, if something like this occurred in the short term, it would be entirely mysterious that the practice went on functioning. But if it took place over generations, it is not an impossible scenario. (Crucially, it does not have to involve the idea of the community being in global error, since we may imagine that the community change their standards gradually over time – like the planks are changed on Neurath's ship). But, this would just be a situation where there are no reliable (moral) belief-forming processes. Hence, there would be no reliability to explain.

not-timelessly true at t_1 – which is an absurdity given the platitude about timelessness of truth. So, if moral truth turns out to be counterfactually mind-dependent in this way, then there appears to be a serious problem for the platonist.

However, the platonist can deny the charge that moral facts are mind-dependent in the counterfactual sense by *rigidifying* the reference to actual moral practices in his account. The minimalist can here take heart from the fact that an analogous move appears to work in other cases – for example the colour-case. We may accept an Euthyphronic account of colours and hold, for instance, that (it is a priori that) x is red just if x appears red to standard observers in standard conditions for observation. We may then solve the problem with counterfactual situations in which the constitution of the standard subjects and the standard conditions for observation is different simply by rigidifying the occurrences of 'standard subjects' and 'standard conditions' in our analysis. Even if one day everybody will become blind, it will still be the case at that time that if actual (read rigidly) standard observers were to see x under actual (also read rigidly) standard conditions for observation, they would deem it red.

By analogy, the minimalist could claim that the existence and character of the set of moral facts depends on the way that actual human beings practice in accordance with moral standards as they actually are – where we should read the occurrence of 'actual' in these claims *rigidly*. That is, if we want to ask what is true about moral issues in a counterfactual world in which there were no, or different, human beings, we would want to ask – exactly as we ask in the actual world – what a well-informed, rational and morally educated subject *in the actual world* (with the subject and moral practice constituted as they are here and now) would judge about that issue.

However, there is a problem with this analogy between colours and morals. Whereas our talk of colours is intimately tied up with our relatively immutable biological perceptual apparatus, our moral talk is merely disciplined by our current, mutable, standards based on our affective natures. We cannot change our perceptual apparatus, but we can (and do) revise our basic moral principles as we change (and hopefully improve) our moral sensibility. So, to rigidify the reference to actual moral practices and human subjects in our defence of mind-independence would illegitimately imply that our actual moral practices were somehow privileged – the

pinnacle of human morality. Our moral language would then – presumptuously – become relativised to our actual moral sensibility.

Let me now outline the way out for the minimalist here. When I just claimed that our set of moral standards is mutable, this was a claim that calls for a qualification. It seems that there are limits to what we could intelligibly count as being moral discourse – and, hence, limits to what moral discourse could change into. To take an example, if someone claimed that the principle 'Lying is wrong' applied to everybody but himself, he could not intelligibly be said to have taken a moral stand. The universalisability of a claim appears to be a necessary condition for its being a moral claim.¹⁵⁹ Other examples of claims that may be regarded as not intelligible as *moral* claims would be this claim: 'Lying is wrong, but I see no reason not to lie'. Or: 'Lying is wrong, but I feel no motivation to avoid lying whatsoever'. The unintelligibility (at least to my mind) of these claims, we may say, is due to the necessary action-guiding and motivation-generating character of moral judgement.

But, if there are immutable principles of morals like this, what are the parts that are mutable then? What we need here, is the distinction between core and peripheral moral principles or platitudes (already drawn in chapter 3). That is, there are platitudes governing the use of moral concepts that are of a very general kind. Platitudes about supervenience and practicality as in the examples above, but also about moral objectivity, substance¹⁶⁰ and general coherentist procedures for moral reasoning, to use Smith's list.¹⁶¹ These are all platitudes that govern how we identify claims as being moral claims in the first place and how we assess moral claims.

But there are, of course, also more specific or substantial moral principles. These are linking-principles of varying degree of generality – from 'Lying is wrong' to 'The right action is the action with the best consequences' – and moral principles that link one moral claim with another – say, 'If lying is wrong then getting someone else to lie is also wrong'. With these first-order moral principles at least some degree of disagreement is intelligible – say between the absolutist Kantian who believes that

¹⁵⁹ I say more about universalisability – and the related notion of supervenience – in chapter 6.

¹⁶⁰ The platitudes about substance are platitudes that help us distinguish moral requirements from other types of requirements – for instance prudential or etiquette requirements. For instance, it is a platitude that 'Right acts are in some way expressive of equal concern and respect'. See Smith (1994), p. 40.

¹⁶¹ Smith (1994), p. 39-41. I have said more about moral platitudes and moral principles – core and peripheral platitudes – in chapter 3.

'Lying is wrong, period' and the duty theorist believing that 'Lying is – other things being morally equal – wrong' or the utilitarian believing that 'Lying is wrong most of the time, since most of the time lying is the least good alternative'. So, we may distinguish two types of platitudes (all of which are, if true, knowable a priori). Those platitudes that are necessarily true – part of what it takes to master moral concepts. And those platitudes that are contingently true – due to the fact that they are grounded in the contingent affective natures of actual human subjects. The former are the core platitudes and the latter the peripheral platitudes.

I now want to suggest that the minimalist may use this distinction to meet Field's mind-dependence challenge. The suggestion is simply to rigidify the core platitudes, but not the more controversial peripheral platitudes – the substantial moral principles. Thus only elements in moral practice that are so entrenched that they are immutable will be rigidified. Hence, when we consider counterfactual claims about morals, we will ask how someone equipped with moral concepts that satisfy the core platitudes would judge – not how someone with our full moral theory, including first-order moral principles, would judge. This would then in a sense ensure – by analogy to the colour-case – that moral facts are not counterfactually mind-dependent.

This solution to the problem about counterfactual mind-dependence also highlights a feature of the minimalist's account of morals that has already been noticed (namely in chapters one and three). If we merely rigidify the core platitudes and ask how a subject with concepts satisfying these core platitudes would judge about moral matters, there is no guarantee a priori that there will be a *determinate* answer. It is not clear on a priori grounds that two subjects with mastery of concepts satisfying the core platitudes, even when we assume that they both have a standard cognitive psychology and use it under standard conditions for observation, would judge in the same way. It is not clear on a priori grounds that there is enough substance in the core platitudes to compel (ideal) subjects to reach a unique verdict on a moral matter. In particular, the affective grounding of moral judgement may leave some matters underdetermined, since the affective natures of subjects may vary with differences in things like culture, sub-culture, individual psychology, generation and gender. By contrast, it seems that in the colour-case, if two subjects are involved in a dispute and master the colour-concepts – and have standard perceptual functions and operate in standard conditions for observation – then it is a priori that there will be a cognitive

shortcoming (ignoring here vague cases) on the part of at least one disputant. So, the minimalist's account of moral discourse must leave room for a phenomenon of moral indeterminacy. And as we saw in earlier chapters, this is exactly the content of the minimalist's contention that moral discourse fails to exhibit Cognitive Command – that there is no a priori guarantee that at least one of two moral disputants has a *cognitive* shortcoming.

4. Minimalism and normativity - rationalism.

As already mentioned, any cognitivist theory of moral judgement must explain two related phenomena: The fact that moral judgement is normative – that is, directed, reason-giving, or action-guiding – and the fact that adopting a moral belief tends to motivate. In other words, the phenomenon that moral belief both points us in a certain direction *and* pushes us in that direction. With respect to normativity, the situation varies according to whether one treats moral facts as *sui generis* or as reducible (in the broad sense that includes both analytical reductions and mere synthetic identities) to natural facts. I now just want to make the point that the platonist as opposed to the Platonist – to complete my comparison of the two – does not have any problems with accounting for the normativity of moral statements.

The platonist's answer, I will argue, is very simple indeed. That is, we may simply take it to be part of the meaning of a moral judgement that it has implications for what we have (justificatory) reason to do or feel. That is, I claim that something like the following is a mere platitude – to take the example of the moral concept of right:

If it is right for agents to φ in circumstances C, then
these agents have (normative) reason to φ in C

Thus, I agree with Michael Smith that a rationalist – in this sense – reading of moral judgement is the right one.¹⁶² Moral facts are simply just facts about what we have (justifying) reasons to do. This is a feature of the discipline that governs our moral concepts. Somebody who claimed on the one hand that φ -ing is right, yet on the other

¹⁶² Smith (1994), p. 60-63.

hand denied that there was (normative) reason to ϕ , would simply not make sense – would fail to recognise something recognition of which is a necessary condition for claiming mastery of the moral concepts involved, i.e. would disregard a core platitude.

5. *Minimalist platonism and explanatory naturalism.*

So far in this chapter, I have tried to show that platonism is both distinguishable from queer Platonism and a viable alternative to ‘bald naturalism’ – viable in the sense that it survives both the ontological and the epistemological queerness-objections. I now want to make the point that there *is* a sense after all in which the minimalist platonist can be a naturalist. Hence, I want to claim that there is a sense in which the minimalist can integrate moral practise in an overall naturalistic conception of the universe. I agree that it is important that morals can *in some way* be seen as compatible with some such naturalistic conception.

The main claim is that there is a basic distinction to be drawn between two ways in which a theoretical treatment of moral discourse may be said to be naturalistic. I will follow Wiggins and Railton¹⁶³ in calling these *explanatory* or *methodological* naturalism and *substantive* naturalism respectively. Crucially, the former is in no way incompatible with minimalism. The explanatory enterprise, I want to argue, can be seen as complementary to – rather than incompatible with – the minimalist's *sui generis* account of moral discourse. Also in the following I want to show how an explanatory naturalism can help the minimalist to explain features of our moral practice – in particular, the truth-aptness and objective pretensions of moral talk. Finally, I discuss the possibility that the minimalist is committed to some kind of moral relativism – hereunder I distinguish many different kinds of relativism.

But, first, let me define the two kinds of naturalism I distinguish. A *substantive* naturalistic theory is a theory that tries to link our moral discourse with non-moral, natural, discourse by identifying moral kinds with non-moral kinds or by showing that moral concepts are equivalent to non-moral concepts – that is, by providing some kind of type-type-reduction. Somebody who takes Jackson's location-problem seriously will exactly want to respond with a substantive naturalist theory.

¹⁶³ Wiggins (1998), p. 351-53, Wiggins (1993a), p. 301-3, and Railton (1993b), p. 315-16.

This can be done in two general ways – I call these analytical and ontological naturalism – depending on whether this link is meant to be discoverable a priori or merely a posteriori. Suffice it to notice now that the moral minimalist – who, as we saw in previous sections, is not moved by Jackson's location-problem – will not be committed to coming up with any such reductive theory. (I shall return to a discussion of the relation between substantive naturalistic and minimalist accounts of morals in chapter six).

A *methodological or explanatory* naturalist, on the other hand, sets out to give an explanatory account of the whole practice of moralising – as seen from the outside, so to speak. He attempts to explain why humans as a matter of fact have developed capacities to adopt normative beliefs, the abilities to discuss with others and to improve these beliefs and the tendency to act on them. These explanations treat moral practice as a natural phenomenon and, accordingly, are formulated in the vocabularies of the various natural or social sciences – like evolutionary biology, sociology or psychology. Thus, there is a sense after all in which the minimalist can live up to the very reasonable demand that he should be able to locate morals in the natural world. But, there is nothing in the explanatory naturalist's approach that forces him to adopt any form of substantive naturalism – though of course he could. He can happily choose to give an account of morals as *sui generis*.¹⁶⁴

Wiggins cites Hume as an example of an early explanatory naturalist who nevertheless sees moral judgements as irreducible and *sui generis*. Hume thought that morals give us 'a productive faculty...' and that '...gilding and staining all natural objects with the colours borrowed from internal sentiment...' it '...raises in a manner a new creation'.¹⁶⁵ With exception of the idea that may be implied in this quote that moral facts are just the projection of 'internal sentiments' that were already there – the picture I have in mind involves rather the idea that culture, and hereunder moral discussion, may modify and refine our 'internal sentiment' – we can agree with Hume. Moral facts are in a sense a new creation – a creation of our educable affective natures and of the practices of argument that we have evolved into being able to participate in. Crucially, the fact that we can give an explanatorily naturalistic account

¹⁶⁴ Wiggins (1993), p. 316 also makes this point.

¹⁶⁵ Hume is quoted in Wiggins, (1993), p. 302. The quote is from Hume (1777), Appendix 1, section 246.

of moral practice does not by itself establish that a substantive naturalistic reduction of the moral to the natural is possible.

The moral platonist sees moral discourse as a practice – a language-game – governed by a norm of truth. He discovers this by taking an analytical approach to the practice. That is, he describes – from within, so to speak – how we actually go about practising in moral discourse and, through this, he arrives at an account of what are the distinctive features of moral discourse. But, he might also have set himself a further task. Human practices or language-games have a point – they have a function in our lives. The platonist might have set himself the explanatory task of seeking out this point or function of moral truth. He would then ask: Why did we humans develop a propensity to accept moral norms and to engage in truth-governed moral debate? In answering this question, the hope is that the features of moral discourse discovered through adopting the analytical perspective – among which are its truth-aptness and degree of objectivity – will be explained.

The general idea is, thus, that the moral platonist, through adding an explanatory account of this kind, is able to locate morality in nature – only in a different sense of ‘locate’ from Jackson's. Let me try to give a short story – a brash conjecture about matters of evolutionary science, to be more precise – to show what such an explanation might look like. Among philosophers who have recently advanced developed theories of an explanatory naturalist kind are Alan Gibbard and Huw Price.¹⁶⁶ I shall draw heavily on their attempts to give explanatory naturalist account of moral discourse and moral truth in the following.

We may follow Gibbard in conjecturing that the human ability to adopt norms has evolved principally because it has given humans a means to co-ordinate action. An ability to co-ordinate behaviour with other members of the species has been a fitness-enhancing feature of individuals. Thus, the ability to adopt norms in animals has evolved in tandem with animals becoming social animals. And, of course, this ability has been enhanced greatly in humans with the evolution of language. That is, we may accept that other social animals have the ability to adopt certain behavioural patterns that co-ordinate their behaviour to social life – to use Gibbard's terminology, we share the ability to *internalise* norms with other social animals. But, what distinguishes

¹⁶⁶ Gibbard (1990), especially chapter 1.4 on 'Normative Psychology' and Price (1988).

humans is an ability to *accept* norms through participation in *normative discussion* – something that is made possible only through language. I shall now say some more to introduce and motivate this terminology and explanatory theory of normative beliefs borrowed from Gibbard.

Let me try to relate Gibbard's notion of *normative discussion* to his distinction between *internalising* and *accepting* norms. An important evolutionary advantage in developing verbal skills is, we may plausibly claim, that it enables humans to exchange information, to co-ordinate reactions to each other and to co-operate in flexible ways. In particular, with the development of language, it becomes possible to discuss with others how to respond to situations other than the actual one – for example, real situations in the past or the future or mere hypothetical situations, perhaps as these occur in fictitious narratives. Through this type of debate, we can plan together how to react to a variety of situations – we can agree on common contingency-plans. Thus, among other things, language allows for the discussion of – and the reaching of a consensus about – how to organise our social life. In *normative discussion*, to define the term, we work out together how we should act and feel in a variety of circumstances – thereby achieving a sophisticated level of co-ordination of action.

Now Gibbard suggests that we distinguish the adoption of norms – and hence the co-ordination of behaviour – through debate with our peers from a less sophisticated way in which we may co-ordinate our behavioural patterns with other individuals of our species. The less sophisticated way is that of merely *internalising* the norms. It is only when we are able – through our linguistic capabilities – to avow and defend the norms that we are governed by in normative debate with our peers that we can be genuinely said to *accept* the norms in question. If the norms merely govern our actions – if we are merely 'in the grip' of certain norms – then we do not accept but have *internalised* the norms. In effect, Gibbard is suggesting that there are no fewer than three motivational systems in play in the human psyche. There is the purely animalistic motivational system – comprising our basic desires like those for food, drink and sex. But, then there is also a distinction between two different forms of normative motivational systems. There is the primitive normative motivational system that co-ordinates action and emotion through the mere internalisation of norms. And,

finally, there is the ‘language-infused system of co-ordination peculiar to human beings’ – that works through the acceptance of norms in normative discussion.¹⁶⁷

However, clearly the mere ability to partake in normative debate and the adoption of norms as a result of this does not ensure that participants co-ordinate their actions and emotions. As Gibbard points out, two further conditions have to be satisfied in order for normative discussion to be a means to co-ordination of action and emotion. Normative debate must *tend towards consensus* and the acceptance of moral beliefs must *influence action and emotion* – or as Gibbard puts it: must have *normative governance*.

First, what is *normative governance*? The benefits of linguistic capabilities – more precisely, normative debate – would not be present if the adoption of beliefs was not coupled with propensities to act accordingly. If, for instance, our shared normative belief that we should allow the chieftain to distribute the food in the group and plan the next hunt did not come with at least some degree of motivation – allowing us to follow the orders of the chieftain rather than our own basic motivations deriving from the animalistic motivational system mentioned above – we would not reap the benefits of the degree of co-ordination and co-operation that this arrangement would result in. But, then it seems that linguistic abilities must have evolved together with a normative motivational control system that *actually influences* the emotions, actions and beliefs of the individuals participating in normative debate. Gibbard calls the feature of acceptance of norms that it tends to influence our emotions, actions and beliefs for the *normative governance* of the norms.¹⁶⁸

Secondly, normative discussion must tend to create consensus – since otherwise the acceptance of normative beliefs would not result in the *co-ordination* of action. But, how does this needed tendency towards convergence of beliefs come about? First of all, according to Gibbard, we pressure each other to be consistent and coherent in the set of norms that we adopt and the particular (occasion-specific) moral verdicts we give. We point to inconsistency or incoherence in the set of norms and

¹⁶⁷ Gibbard (1990), p. 75, see also pp. 56f. Interestingly, Gibbard argues that the Milgram experiments – where individuals in certain famous experimental settings are lured into acting in ways that conflict with their *accepted* normative beliefs because they are *in the grip of* conflicting norms – are examples of conflicts between these two normative motivational systems.

¹⁶⁸ Gibbard (1990), p. 72.

particular moral verdicts (about real or hypothetical situations) that other disputants avow – and we 'discuss with ourselves in our heads' to root out incoherence in our own system to prepare for real normative discussions. But, this is not enough to ensure that disputants will converge. Thus, secondly, in order for there to be a tendency towards convergence in norms, we have to be to some degree persuadable. Or rather, whereas, on the one hand, we do have some need for firmness in our moral stands – we cannot be in a perpetually undecided state such that we never actually accept any norms and act on them and we cannot be total pushovers in normative debate – we must, on the other hand, also to some degree have an innate desire to converge with our peers. This tendency, together with the leverage we get on others through discovering incoherence in their set of norms, explains why we often manage to converge.

In short, I hope here to have drawn a plausible sketch of how we – with Gibbard – can give a methodologically naturalistic explanation of the evolution of normative debate and normative belief-systems ultimately by reference to the fitness-enhancing character of an ability to co-ordinate action and emotions with other individuals.

6. Explaining the usefulness of a concept of truth in morals.

So far, then, we have outlined a way in which evolutionary sense might be made of why humans have become believers in common norms and why they engage in normative debate. Now, the notions of acceptance and normative debate – and in particular the notion of consistency – as used in the foregoing have already implicitly introduced the notion of truth. Where it makes sense to demand genuine consistency there must be truth-apptitude. But, exactly where and how does truth enter the picture? How might the adoption of a standard of truth have helped? Huw Price has, I think, good answers to these questions in his general explanatory theory of truth, so let me now follow his lead for a while.¹⁶⁹

The short explanation of why we have come to apply a notion of truth in normative debate is that through becoming seekers of a common truth we have come

¹⁶⁹ Price (1988), chapter 7.

to dislike disagreement and value convergence and reasoned argument – and the latter has as a matter of fact been an evolutionary advantage for us.

First of all, we must here understand the conceptual connection between truth-aptness, on the one hand, and disagreement, intolerance and the emergence of reasoned debate on the other hand. As we shall see, the applicability of a standard of truth to a certain subject-matter makes disagreement about the subject socially unstable and creates an incentive to seek a common point of view through reasoned debate. And – vice versa – the fact that disagreement within a discourse is socially unstable and can be resolved through reasoned debate makes the discourse apt for truth.

If I believe that p and you believe that not- p then we cannot both have true beliefs about p . If I seek the truth of the matter, then I want my belief to have the award of being true. Therefore, I will, at least in circumstances that allow free debate, want to publicly challenge your belief. I will think that you are at fault and I will want to show that I can back up my belief better than you can back up yours. The adoption of a common standard of truth, therefore, becomes an incentive to engage in reasoned argument – hereunder to share all relevant information about the matter – whenever disagreement is discovered and thereby it promotes a convergence in beliefs. The result of applying a concept of truth, therefore, is the creation of reasoned argument and the development of an intolerance of disagreement. On the other hand, the possibility of reasoned argument and the public intolerance of disagreement seems to be exactly the degree of discipline that is sufficient – together with the syntactical requirement – for a discourse's being truth-apt (as was already pointed out in chapter one).

So the possibility of our use of a standard of truth in a discourse is conceptually linked (bi-conditionally) to the possibility of reasoned debate and a tendency towards reason-based consensus. But, to explain why a given discourse is truth-apt, we must then ask the question: Why has reasoned argument and the incentive to reach a reason-based consensus – the intolerance of disagreement – about the subject-matter of ethics been beneficial to our species? The answer here cannot, of course, be that disagreement indicates that one or the other disputant has a *false* belief. This would be to presuppose the existence of the standard of truth that we are trying to explain. So, another explanation must be found.

Now, as Price convincingly argues, reason-based convergence and reasoned debate about a given subject-matter is beneficial for the species under a certain assumption – namely the assumption that beliefs about the subject-matter have *similar behavioural consequences for individuals that hold them across a community*. The emergence of a reasoned, truth-governed debate over a given subject-matter can, thus, be explained by showing that the assumption *is* satisfied. And, of course, it may be the case that the explanation of this differs from discourse to discourse. Let me try to spell this out in more detail for the remainder of this section.

As Price notices, the utterance of the same sentence (in the same context) by different individuals in a community tend to be associated with mental states that have similar behavioural effects for the individuals. Therefore, disagreement between two individuals – their having contradicting beliefs – tend to mean that they differ with respect to behavioural dispositions as well.¹⁷⁰ So, the question above – about why reasoned consensus is sought after – can be reformulated as: Why do we value sharing behavioural dispositions with respect to a given subject-matter? Here Price's answer is that creating convergence in beliefs – hence sharing behavioural dispositions – is beneficial under a certain assumption. Namely, that we are all *in the same boat* with respect to these behavioural disposition, to use Price's useful expression – that these behavioural dispositions are *equally beneficial* to all individuals across the community. He defines the Same Boat Property as:

‘A class of mental states have the Same Boat Property if their typical behavioural consequences are such that their behavioural appropriateness, or utility, is predominately similar across a speech-community. If a mental state has the SBP, then if it is appropriate for any one of us, it is appropriate for all – we are all in the same boat.’¹⁷¹

So, with Price, we end up with the following observations. Truth-aptness is linked conceptually to the possibility of reasoned argument and to intolerance of disagreement. To explain why truth-governed debate among humans has emerged

¹⁷⁰ Of course, not all disagreements in belief actually result in different behaviour on all occasions. Beliefs and desires explain actions holistically. Differences in belief can be compensated by differences in desire such that same behaviour results on particular occasions. If you believe that the next lecture in the lecture-room will be on Heidegger and like to hear about his philosophy and if I think that the next lecture will be on Wittgenstein and like his philosophy, we will both head for the lecture-room despite our divergent beliefs. The following argument is based on the plausible assumption that scenarios such as this are relatively rare.

about a certain subject-matter, we must explain why adopting a standard of truth has been beneficial to humans. And that, in turn, must be the same as explaining why reasoned argument and intolerance of disagreement has been beneficial. But, as it turns out, the latter can only be explained under a certain assumption. Namely that having a given belief about the subject-matter in question has the *same* benefits for individuals *across the linguistic community* – that we are all in the same boat. Thus, truth-apt discourse and reasoned argument about a given subject can be explained only if beliefs about the subject has the Same Boat Property.

So, what sets of beliefs have the Same Boat Property – and why? In particular, does *moral discourse* have the property and, if so, why? And, following closely on this, can we explain the truth-aptitude of moral discourse from a methodologically naturalistic point of view? An important initial point to make is that even among the discourses that do have the Same Boat Property there may be *different* explanations of why this is the case. An external standard of truth – and the reasoned arguments that it drives us into – may have different functions in different discourses. Take first the example of discourse about our immediate natural surroundings. Here all humans are clearly in the same boat for the simple reason that we share these surroundings and, by and large, we have the same cognitive apparatus, the same basic needs, vulnerabilities etc. To take an example, we are all in the same boat with respect to believing that a certain fruit is poisonous or that there is more game in the next valley. It is, therefore, an advantage for us to *share information* and *think together* about such things. When we pool our information and we co-operate on processing this information – detecting informational and rational error in each other's belief-systems – we are bound to acquire beliefs that are better adjusted to our surroundings. We benefit from the experiences – the wisdom – of our whole community. Partaking in the ensuing informed consensus will benefit us, since the benefits of the associated behavioural dispositions are the same for all of us.

But, what about moral discourse? Is there a sense in which we are all morally in the same boat? I shall argue that there is, but also that it is not the same sense as in the case of beliefs about our natural surroundings. We benefit from trying to achieve convergence in our beliefs about our natural surroundings because we are trying to

¹⁷¹ Price (1988), p. 152.

represent the *same* reality in our individual belief-systems and consensus is a reliable indicator of successful representation (by the motto that many think and observe better than one). But, the idea that moral discourse is about some independently existing moral reality – the Platonist's idea – that we can benefit from exploring together is, as already argued, simply not attractive. It would seem that the direction of fit is opposite in the case of morals. Through truth-governed argument we *create* a certain consensus in our community. We seek moral consensus – that is, we value argument and truth in moral discourse – because consensus in moral beliefs means added benefits from social co-operation as I argued with Gibbard in the last section. For instance, agreed norms for distributing the goods prevents strife and creates a security-network – to the benefit of most in the community. The function of moral discourse is not to represent a pre-existing reality, but to regulate and possibly improve our communal life through the achievement of and the process of achieving normative consensus.¹⁷²

Let me be more specific about what the benefits of moral consensus in a community are by thinking through an example. I shall, to anticipate, argue that there are two distinct ways in which moral argument and consensus can be beneficial – through providing co-ordination and through adjusting the norms of a community to the external conditions of the community. Imagine that a small, closed, community – a village or perhaps a beduin-tribe – centuries ago adopted a set of norms strictly prohibiting pre-marital sex. Let us say that this made a lot of sense in that community. The small size of the community and the way it depended on close co-operation for its very survival made the prevention of dissent and strife of the utmost importance. And premarital sex – and especially the unwanted pregnancies – had a strong tendency to create jealousy and animosity in the community. In addition, it tended to disrupt the complex family-structures and, in particular, the functioning of the fine-tuned inheritance-laws that played an important part in ensuring that property was distributed in such a way that all members of the community could survive. The internalisation of the norm was especially beneficial because when couples broke the

¹⁷² Notice a disanalogy between the role and value of consensus in moral discourse as opposed to discourse about our immediate surroundings. In the latter, consensus itself has little value. What is important is the process leading us to consensus – the pooling of evidence etc. And, it is of course important that disputants do not seek consensus for its own sake, but truth for its own sake. In

code (which they sometimes did), they could only do so with a strong feeling of guilt. (The norm being therefore both grounded in *and* normative for these feelings). This, in turn, ensured that they would not reciprocate the anger of the other villagers when found out. There would then be no open and highly disadvantageous strife, but merely a ritual acceptance of the moral indignation of the community through a display of remorse and perhaps payment of damages. And, eventually, there would be a re-entry in the community. So, all in all, the internalisation of the strict code against premarital sex was a good thing in that society. The degree of co-ordination it facilitated made for a much more peaceful and co-operative, hence successful, community. On average, it promoted the survival and the well-being of the individuals of the community. Thus, I have here given an example of the benefits of a moral consensus deriving from *co-ordination*.

However, it may now be objected that this story does not show why the normative consensus about premarital sex *had* to be imposed through normative debate. All that was needed was *some* way of imposing the code. And, indeed, in many communities of the kind used in the example the code would be imposed through authority – often religious – rather than through debate and acceptance (in Gibbard’s technical sense). And this is a quite general point. Co-ordination can be brought about by somebody telling us all what to do – by the forced imposing of dogmas. So, normative debate is one way of bringing about co-ordination, but not the only one. It is a sufficient, but not a necessary condition for achieving co-ordination. Though it may be the most efficient means, as I shall return to in the following.

To see why normative debate – including the adoption of an external standard of moral truth – is useful in a second way, and better than deference to authority as a means to organise society, let me continue my story. Let us now say that the community develops over the centuries into a society much like ours today. The villages become integrated into larger communities with bigger cities, family-structures become less important as ways of distributing property and creating social security, survival becomes easy due to more efficient means of production, people tend to marry later due to the need for education, contraception becomes more efficient, abortion less risky etc. Now, most of the causes behind the general

morals, on the other hand, consensus is itself a goal, not merely the pooling of information –

acceptance of the norm prohibiting premarital sex have disappeared. And we may imagine that there comes a time where the cost of accepting the norm – the pain of guilty feelings and sheer loss of opportunity for pleasure and experience – becomes higher than the gains.

The point is thus that normative debate may help a transition to a more timely moral code. If the code had been imposed by authority and dogma such a transition to a better set of norms – by the standards of the community itself – may have been difficult to effect (at least without revolution and strife). By contrast, if public normative debate was possible, people might raise the question of why a given moral code is adopted – they may demand that justification is offered. In the course of the ensuing debate, new suggestions for ways of regulating life in society can be advanced and could end up winning the debate.

Now, just as with the first benefit of normative debate and normative consensus – namely the creation of co-ordination – it may be objected that it is possible for one person – perhaps a religious authority-figure – to bring about the needed adjustments in the moral code. Thus, it is not strictly speaking necessary for there to be *debate* over the issues – understood here as the exchange of arguments between different persons. For instance, one might imagine that the Pope alone was assigned the role of being the one who thought about how to adjust the norms of the society to the changing circumstances – to create the needed dynamism in the norms of the community.

However, this is no serious objection to the above point. The point of the foregoing is that normative debate brings benefits because it opens the possibility for reasoned argument and thereby the taking into account new circumstances. And reasoned argument does not have to be interpersonal. Even the Pope has to ‘discuss with himself’ – engage in silent reasoned argument – in order to reach a *rational* decision. (If he did not engage in at least silent reasoned argument, his conclusions would be arbitrary and therefore highly unlikely to improve the moral code in the light of the changing circumstances). So, reasoned debate would have to be *possible* in this scenario even if not practised.

although, again, properly taking part in moral debate is seeking truth, not consensus.

Anyway, another important point helps the case for public normative debate as opposed to an authority-imposed moral code. Though it is *possible* to arrange a society in this way and let normative matters be decided by authority, it simply seems odd to let the judgement of only *one* person help in deciding how to adjust the norms of the society. Surely, here as well as everywhere else, the principle that many minds think and observe better than one applies.

Thus, a second benefit of truth-governed normative debate is the efficient achievement of some degree of dynamism in the set of norms adopted by a community. The provision of an ability to respond to changes in the circumstances of the community and a chance to defend new, more timely, ideas about how to organise life in the society. Normative debate can create opportunities for *improvements and adjustments* in the set of norms – in addition to the already discussed advantages of co-ordination.¹⁷³

To sum up, the usefulness of adopting an external truth-standard – and hence the explanation of why it is adopted – varies from discourse to discourse. It is always necessary to make the assumption that all participants in the discourse are in the same boat. In the moral case, in particular, we all benefit from treating each other *as if we were in the same boat*. Doing this enables us to reap the benefits of *increased co-ordination* due to the achievement of consensus and *improvements* in our system of norms due to a dynamic responsiveness to changes in our social and natural environment.

Ending now my excursion into naturalistic explanation, let me note that what has been provided in the foregoing sections has been the beginnings of a (methodologically) naturalistic explanation of why we have truth-governed *moral* debate in the first place. Perhaps we can also now glean a naturalistic explanation of why the truth-predicate governing different discourses have *different* realism-relevant properties, since we have now discovered differences in the function of truth in

¹⁷³ An important part of this story of changes in moral codes is the fact that wholly *new* problems – and solutions – occur as technology changes. (In the example above, the invention of efficient contraception and less risky abortion called for a change in normative belief). Examples from current moral debate are, of course, the development of gene technology and the new threats to our existence due to environmental damage. Here we face choices and problems that no earlier generation faced before us and we have the problem of extending or changing the moral code that was handed down to us to deal with the new cases. Truth-governed normative debate can function as a means to create a consensus about such new questions.

different domains. (Thus, maybe the minimalist can complement his set of analytical distinctions between discourses with respect to their objectivity with a naturalistic explanation of why there are differences in robustness of subject-matters – why the properties of the norms of truth in different domains differ). However, the main point of the foregoing has been to show how the minimalist is able to ‘locate morals in nature’, not through a substantive naturalistic account of morals as Jackson would have wanted, but through giving a methodologically naturalist account of moral practice along the lines laid out in the foregoing. The contention is thus that the moral minimalist can avoid any suggestion that his platonism is super-naturalistic and therefore ontologically queer. In short, it has been suggested that one does not have to be a substantive naturalist in order to avoid the charge of being an anti-naturalist.

7. *Forms of moral relativism.*

Having seen how a naturalistic explanation of why moral discourse is truth-apt could go, I now want to finish this chapter with a closer look at minimalism in relation to relativism. Mainly I want to argue that the explanatory account of morals – that is, an understanding of the *function* of truth in morals – can show why the moral minimalist has a dialectical advantage in relation to (certain kinds of) moral relativism. I want, finally, to argue that the naturalistic account of the truth-aptness of morals squares with moral minimalism – and that minimalism can be regarded as a new, viable, form of moral relativism.

Before going on with the comparison between relativism and minimalism, let me, however, clarify some terminology and make some observations about various *kinds* of relativism. Price suggests that we may see relativism in the light of his discussion of the function of truth as arising from the situations in which both disputants realise that – despite the fact that there appears to be a dispute – neither is mistaken because they realise that they are no longer in the same boat. Price calls this general phenomenon *the evaporation of disagreement*. However, this discovery – that neither disputant is mistaken – can take at least two forms according to Price: 1) The evaporated disagreement may be an *evidential* dispute or 2) it may be an *inferential* dispute. As we shall see, these kinds of evaporation give rise to two different types of relativism and I shall, therefore, discuss two ways of cashing out the idea of

evaporative disagreement – namely weak relativism and parochialism (i.e. strong relativism as, for instance, defended by Harman).

More precisely, I shall argue that it is only the latter type of relativism – parochialism – that is at odds with minimalism. That is, if we agree that evaporation of inferential disagreement is possible within moral discourse, then what I shall name ‘relativistic practice’ is sanctioned.¹⁷⁴ By ‘relativistic practice’ I shall understand a certain way of dealing with *seemingly* irresolvable moral disputes. The way, that is, that simply consists in concluding that there *is* no objective way of settling the dispute and of then saying that it is simply up to the individuals to decide ‘what is true for them’. Thus, an advocate of relativistic practice within moral discourse would hold that there are situations where it makes sense to simply back down in the dispute and tolerate the other disputants point of view – to simply give up trying to live up to an external standard of truth and to give up trying to resolve differences through normative discussion.

First of all, let me look at a harmless sense in which disputes might evaporate – what Price calls evaporative disagreement on evidential grounds¹⁷⁵ – and the ensuing weak and harmless form of relativism. The example above – about the community with norms prohibiting pre-marital sex – can provide us with an example. The way I told the story, I gave the impression that the strict code was appropriate in the context of the early closed society and became inappropriate later in the modern context. Thus, I did not claim that someone from the old society would strictly speaking disagree with someone from the modern society even though they accepted different everyday norms. They *did* differ with respect to the everyday norms they lived by, but not necessarily with respect to the underlying rationale of their moral code. That is, they might agree on certain basic linking-principles – norms of a higher order than the everyday norms – and their ‘disagreement’ may simply be due to the fact that they were operating with different (but compatible) evidence about the different societies in which they were living.

Say that both agreed on the following two principles: P₁: In a society with socio-economic structure S₁, adopt norm N₁. And P₂: In a society with socio-

¹⁷⁴ I borrow the term from Price (1988), p. 178f.

¹⁷⁵ Price (1988), p. 173, 178. This position is what Gibbard calls ‘relativism’ in his (1990), p. 205, 208-11.

economic structure S_2 , adopt norm N_2 . (Where N_1 and N_2 are implicitly contradictory). And say that they believed that their society had structure S_1 and S_2 respectively and detached appropriately. Once they discovered that they were living in different types of society, their disagreement about norms N_1 and N_2 would simply disappear – they would see no mistake in each-other's views.¹⁷⁶ But, of course, this is a very weak form of relativism indeed. Crucially, this point of view is not relativistic all the way down, since the higher order norms are treated as absolute and there is no reason to accept relativistic practice in moral discourse. As such it is compatible with minimalism and even stronger forms of moral objectivity.

I turn now to the much more disturbing form of relativism that appears if we imagine that evaporative disagreements are possible about basic linking-principles like P_1 and P_2 – what Price calls *inferential* disputes. An inferential dispute is what we would understand as a basic moral disagreement – a dispute about a basic moral (linking-) principle. But the idea of an evaporating basic moral dispute *does* imply that relativistic practice is acceptable. That is, if we can we countenance scenarios of moral disagreement in which – upon our discovery that we do not share basic moral principles – we are simply allowed to back down and accept that neither party to the dispute is mistaken, we accept that our moral views do not have to be taken seriously by others who do not share them. Ultimately, we would have to *tolerate* the basic moral points of view of these others and accept that no-one has an obligation to justify his views in normative debate with others of a different basic moral persuasion.

Gibbard calls any view, implying that the evaporation of inferential disputes is possible, *parochialism* – what many would call relativism proper.¹⁷⁷ Basically the parochialist moralises in such a way that he simply excludes the opinions and arguments of certain others. And crucially not because he can tell some story of why these others are epistemologically impaired in some way. In fact, he might recognise that the outsiders have all the relevant non-moral information and that they reason correctly and still – according to the moral parochialist – he may intelligibly refuse to take their point of view into account. In other words, the only answer the parochialist

¹⁷⁶ Gibbard (1990), p. 208-11, suggests an example of what the underlying basic norm of the relativist might be – namely, 'Accept the ethos of your local society'. That is, the relativist might hold that we have a basic 'communitarian commitment' to accept whatever consensus about everyday norms has emerged in the normative debate in our local community.

can give when asked why he does not take the views of the person from another group into account is simply: 'He's from that other group!'

In short, there is, according to the parochialist, no objective truth about moral issues and relativistic moral practice is sanctioned.

8. Moral relativism and the function of moral truth.

What I now want to argue is that parochialism about morals cannot be squared with the explanatory account of the function of moral debate – in particular of our use of an external standard of truth in moral discourse – given above. I shall take as an example Gilbert Harman's most recent formulation of parochialism – i.e. strong relativism.¹⁷⁸ Ultimately, I want to support the claim that this gives the minimalist a dialectical advantage over the parochialist.

Harman tells us the following story to show why we should become relativists about morals. He claims that we have inherited a non-relativistic moral vocabulary that presupposes that there is a single true morality. But, now we (ought to) have discovered that this presupposition is false – as the nihilist or error-theorist also believes themselves to have discovered – as we have become acquainted with many different conflicting moral frameworks of which we have realised that no single one can be held to be privileged.¹⁷⁹ Thus, to use the vocabulary from above, according to Harman we have learnt that morals must be a parochial matter.

But, Harman does not want at this point to draw the conclusion that moral discourse should be abandoned altogether. He draws an analogy with Einstein's discovery of the relativity of mass to a spatio-temporal framework. Our inherited notion of mass treats mass as something absolute – something relative to one

¹⁷⁷ Gibbard (1990). p. 205-8. Gibbard himself admits that parochialism is what philosophers normally call relativism (see p. 205n1).

¹⁷⁸ Harman and Thompson (1996).

¹⁷⁹ Thus, Harman's argument for relativism depends at bottom on an argument for a presuppositional error in all moralising. This makes his position vulnerable to the criticism that we do not know for sure that there is no single true morality. No matter how much actual radical disagreement there is this may all be down to a mere temporary failure to discover ways of resolving moral disputes – a failure that may be remedied some day. (As argued in previous chapters, all we can legitimately claim is that we, so far, must remain agnostic about whether all moral disputes involve cognitive shortcomings – and hence are solvable). Or, it may be that the multitude of moralities are really just applications of higher order, absolute norms to different circumstances – as the weak relativist above argued. So, we do not even have to concede Harman the first steps in his argument.

privileged spatio-temporal framework. So, strictly speaking everyday mass-discourse from before Einstein till our day rests on a false presupposition. Harman's point is that we did not have to abandon mass-discourse after Einstein. We can make sense of everyday mass-discourse by ascribing relativised truth-conditions to statements within it. So, 'x has mass M' is true as uttered by a speaker if and only if x has mass M in relation to the speaker's spatio-temporal system.

Now, Harman asks, why not do something similar to make sense of ethical discourse after our discovery of the multiplicity of moral frameworks and the failure to single *one* out as privileged? He suggests that we ascribe relativistic truth-conditions to a moral statement with the following words:

'For the purposes of assigning *objective* truth conditions, a judgement of the form, *it would be morally wrong of P to D*, has to be understood as elliptical for a judgement of the form, *in relation to moral framework M, it would be morally wrong of P to D*'.¹⁸⁰

(Different versions of this view can of course be constructed depending on whether the moral framework spoken of is the moral framework of the speaker or of the speaker and his audience or perhaps of the speaker's culture as a whole). Now, this view of moral discourse would not make all moral debate impossible, since one could still have debates within one's parish. Thus, relativism of this kind does not fall prey to a common objection to relativism, namely that it leaves no room for a distinction between seeming right and being right. As Harman points out a moral framework should be thought of as a 'system of *corrected* values'. It is not just the set of principles I/we happen to have now, but a consistent system of principles constructed from the set I/we have now. In other words, I/we may be wrong about what my/our moral framework says and the 'seems right'/'is right' distinction is saved for the relativist.

Also, Harman's relativist does not – immediately at least – have trouble with the straight-forward inconsistency of the crude relativist who simply claims that both parties to radical disputes have true, but contradictory, beliefs. That is, by relativising the contents of moral claims to frameworks, there are no radical disagreements of the form: A believes that *p* and B believes that not-*p*. Hence, it is not a consequence of

¹⁸⁰ Harman & Thompson (1996), p. 43.

this form of relativism that claiming the truth of the views on both sides of a radical dispute involves claiming that p and not- p .

The problem for Harman's parochialist comes only when he has to deal with people from other moral parishes – people with other moral frameworks. On Harman's account, the moment the disputants discover that they do not share moral frameworks, their disagreement evaporates for the simple reason that they realise that they are not speaking the same language. According to Harman's strong relativist, if A says that 'It is right to ϕ in C' and B denies this – and if A and B do not share moral frameworks – then what A is saying is really that 'According to my moral framework, it is right to ϕ in C' and what B is denying is that 'According to my moral framework, it is right to ϕ in C'. Thus, the univocality of cross-framework disagreement disappears, since the relativised contents of the opposing views no longer contradict each other. Every deep moral disagreement must evaporate on Harman's parochialist account.

Harman, being a proud relativist, embraces this fully and accepts the consequence that rational moral debate – seen as a process aiming at framework-independent truth – must be abandoned. But, notice that it then would seem that he advocates giving up the tool for social co-ordination and adjustment of moral norms that, as was argued by the explanatory naturalist above, truth-governed moral debate is. The discussion of the function of moral truth explained exactly why relativistic moral practice as advocated by Harman is not the way we *actually* practice in moral discourse. We need normative debate in order to discuss – and perhaps create consensus – even with people that do not currently share our basic moral outlook (at least if they are people with which we have to deal). So, it would seem that Harman's relativistic substitute simply cannot do the job of *real* moral discourse. There is, it seems, a tension between his account of morals and what we know about the function of morals.

But this may be too quick. Harman believes that he has something to put in stead of a (non-relative) standard of moral truth to do the job of creating consensus. He suggests that moral discourse across frameworks should instead simply be seen as 'moral bargaining' – a process similar to bargaining over the price of a house.¹⁸¹

¹⁸¹ Harman & Thompson (1996), p. 20-31.

According to this view, the consensus that is the outcome of moral 'discourse' is a compromise and its exact shape will depend on the relative power of the parties taking part in the bargaining rather than the relative strengths of their arguments. According to Harman, we simply ought to accept this and thus practise morals with a relativistic self-understanding.

The question is, however, whether moral bargaining can be an adequate substitute for truth-governed moral debate as a means to co-ordinate social life and improve our moral beliefs. Let me finish my discussion of Harman's parochialism with an argument why it cannot. If we bargained with people from other moral parishes about what basic moral principles to accept, then we would have to be conscious about the fact that the norms, we accept, do not have any objective aspirations, but are mere 'bargaining-positions' masking the personal (or group) interests of the disputants. But, then the question is whether moral beliefs could have the normative governance – the effect on motivation and action – that the adoption of moral norms actually have. In other words, can Harman's relativist account for the *categorical force* of moral beliefs – the feature of moral beliefs that they call for a certain course of action no matter what other, conflicting, desires the agent has.

Let me first try to make the point by considering an example – namely the well-known scenario of the Prisoner's Dilemma. This example can illustrate how the adoption of norms by a group of individuals – resulting in the co-ordination of their behaviour – can bring benefits to the group. Thus, if both prisoners accept a norm against ratting and see this norm as having a claim to truth – take it seriously, so to speak – then their acceptance of the norm will tend to motivate them accordingly and, (if it motivates to a sufficient degree) they will both keep silent. Hence, in the given scenario, they will reap the benefits of co-operation.

However, if the prisoners see the norm against ratting they accept as a mere bargaining-position in a power-struggle – if they have the relativistic self-understanding that Harman recommends – there would appear to be little to stop them from disregarding it in their present circumstances. The relativistic attitude towards the norm would undermine its normativity and motivating effect. The prisoners would have to adopt a form of *neutral* or *external* standpoint with respect to their own moral beliefs – by seeing them as mere bargaining positions that they happen to hold – and therefore they would not have reason to take them seriously as

action-guiding beliefs. The normative beliefs would not discourage them from doing what would appear to them to be *now* in their best subjective interest – namely to rat. Hence, the adoption of the relativistic attitude towards their own normative beliefs would ultimately deprive them of the benefits of co-operation. Contrasting this with the minimalist's account, there appears to be no problem of this kind for him. The affective grounding of entitlements to moral beliefs, on this account, ensures that the beliefs – other things being equal – are reason-giving and motivational for the agent.

We can now see more clearly what is wrong with Harman's and other forms of parochialism. First of all, it seems that the relativist's account cannot help us to naturalistically explain why truth-governed moral discourse has emerged in the first place. Since, on his account, it does not seem that the benefits of truth-governed moral discourse – the benefits of co-ordination and adjustment – could have come about.

But, secondly, if we probe a little deeper, we can see that the parochialist's problem here is really one of reductionism. What Harman suggests is that we accept that moral content can be analysed in wholly non-normative terms. He thinks that 'It is wrong to ϕ ' can be analysed as 'According to my moral framework, it is wrong to ϕ '. That is, in terms describing some relation between the set of moral principles that we adopt and a moral judgement. The suggestion is that we can step out of our normative perspective – detach ourselves from our own affective natures – and adopt a morally neutral point of view in relation to which every moral judgement is relativised and thus also a stance from which there is no way of *morally* ranking the different moral frameworks. But, then it is no wonder that – for people who come to believe in this possibility of a reductive analysis – the moral judgement will no longer have any normative and motivating force. Since it is simply no longer seen as a *normative* judgement in the final analysis.¹⁸²

Let me, however, attempt to challenge the general conclusion – that parochialism with its advocacy of relativistic practice is at odds with an adequate explanation of the function of truth in morals – one final time. Perhaps we have overlooked one thing. The explanatory account given above explains why we have needed truth-governed debate – namely primarily to reach normative consensus and

¹⁸² A similar criticism of the parochialist, I believe, is found in Johnston (1989), p. 166-69.

co-ordination. But, what about disagreements with people that we do not actually have to have dealings with? Price makes the following point in passing:

'The advantages of conformity may sometimes be outweighed by the disadvantages of intolerance. 'Live and let live' may sometimes be the optimal survival strategy. Perhaps we should behave as if everyone in our own community is in the same boat, but allow that foreigners do things differently'.¹⁸³

Thus, an argument might be given for the claim that, since I am only interested in co-ordinating my normative beliefs with other members of my community – those with whom I interact and on whom I depend – I am in a sense only interested in treating other members of my own community as in the same boat as me. Hence, it is not clear that I have to disallow evaporative disagreements with representatives of communities that are far removed from our community – whether geographically or historically – in order for moral discourse to fulfil its purpose.¹⁸⁴ So, is there after all a way in which the function of moral debate sanctions evaporative disagreement *at the fringes*, so to speak?

I think there are several points to make against this suggestion. First of all, it should be kept in mind that I am never allowed to suspend moral judgement in any *actual* moral dispute, since presumably anyone with whom I am in a position to argue would be someone that I am in some sense in contact with and whose actions might affect the welfare of people in my community. Arguably, the increase in global economic and ecological inter-dependence makes 'our community' into a global one. So, in practice, this form of relativism does not license any degree of moral tolerance or suspension of moral judgement.

Secondly, it should be realised that the argument suggests only the possibility of suspension of moral judgement about faraway *people*, not the *moral views* of the people from faraway. Intuitively, there *is* something odd about criticising people from the past or remote places who merely conform to the norms of their society and time.¹⁸⁵ It is in some sense intuitively inappropriate for me to criticise, for example, the Vikings for adopting the norms they adopted and the way they acted. I would

¹⁸³ Price (1988), p. 180.

¹⁸⁴ This may be the point of Bernard Williams' 'relativism of distance' – his claim in particular that some ethical confrontations are merely 'notional'. See Williams (1985), p. 162.

suggest that this intuition comes from a deeply held view. That is, we tend to accept and even value to some extent that people conform to the norms that are the generally accepted norms of their time and place. Perhaps exactly for the reason that part of the function of normative discourse is to establish a normative consensus. And we may do this even in cases where we – from the perspective that our set of (hopefully) more developed norms now give us – find the old consensus morally unjustifiable. We may thus accept that we ought not to criticise and condemn the Vikings – the actual group of deceased individuals.

On the other hand, accepting that we ought to suspend judgement in cases like these does not result in the advocacy of relativistic practice. Thus, certainly, if anyone attempted to revive the norms of the Viking society – to advance it in moral debate as an option for us today – we would have to disagree with him given that our moral standards are radically different from those of the Vikings. We do not give up a claim to objective truth. We only, in special cases, accept that certain people had no chance of knowing any better and that they were therefore excused that they had bad morals.

Thirdly, the quote from Price suggests that it is sometimes a better survival strategy not to raise normative issues with others – to have a 'live and let live' attitude. This point need not be a point in favour of the possibility of evaporative disagreements, though. I may disagree vehemently with someone on some ethical matter and yet choose not to voice my dissent for lots of prudential reasons. I might, for instance, be afraid of the Inquisition or the Secret Police and therefore choose to keep my moral outrage to myself. Or I might just not bother to challenge the person I disagree with – perhaps because I know that he is too dogmatic to listen to my arguments anyway. Thus, it may be better for me not to start up a normative discussion in certain circumstances, but that is not allowing for evaporative disagreement in the relevant sense. I still disagree, but just prudently choose to keep quiet about it.

Finally, there is a sense in which we can even understand the usefulness – and hence explain the existence of – a moral practice that takes any other moral framework seriously no matter how distant its defenders. That is, a practice that disallows any evaporation of (inferential) disagreement whatsoever. It makes sense,

¹⁸⁵ Again, this intuition seems to be what Williams (1985) is drawing attention to.

that is, to take any challenge to one's system of beliefs seriously because testing the justification and coherence for one's beliefs is always potentially a chance to improve one's system of beliefs. Maybe the foreigners are not people we have to deal with in our everyday lives, but their views about how we ought to organise society might nevertheless prove useful to us. Thus, it does make sense to take the attitude to others that Gibbard calls 'widening communities' – the anti-parochialist, absolutist, attitude treating any other person (thinking and acting being) as qualified to be one's opponent in normative debate and as a potential winner of this debate.¹⁸⁶

9. Moral minimalism and the function of truth in morals.

So, parochialism – and the relativistic practice in moral discourse that it results in – is a very problematic view. It seems to both be at odds with an explanatory account of the function of truth-governed moral discourse and to involve a form of reductionism that leaves it a mystery how the moral beliefs of someone with a parochialist self-understanding can have categorical force. What we have seen is rather that an understanding of the function of moral discourse leads us to expect that moral *practice* is in a sense absolutist. To turn finally to the minimalist, I will now briefly argue that the minimalist does not have similar problems. The explanatory account of moral practice given coheres well with the minimalist's analytical account of morals. Therefore, the minimalist appears to have a clear dialectical advantage over, for example, Harman's parochialist.

There are two central features of the minimalist account. To understand the relation between minimalism and relativism we must briefly remind ourselves of these. First of all, there are the basic claims that morals is truth-apt and that moral truth is epistemically constrained. And, secondly, there is the claim that moral discourse fails to exhibit Cognitive Command. As we saw, this amounted to the adoption of an agnostic attitude with respect to whether it is a priori that moral disagreements involve cognitive shortcomings.

I now want to point out that it is implicit in the minimalist position that relativistic moral *practice* must be rejected. Relativistic practice involves a certain way

¹⁸⁶ Gibbard (1990), p. 211.

of dealing with *seemingly* irresolvable moral disputes. As already mentioned, advocating relativistic practice involves recommending – in the face of seemingly intractable disputes, of disputants having different basic moral views – that we conclude that there *is* no objective way of settling the dispute. And then it is proposed that we see our moral views as implicitly relativised. But, a rejection of relativistic practice of this kind was implicit in the minimalist’s claim that we cannot accept that there are disputes that are absolutely un-settleable. Remember, that is, that the minimalist had to avoid Scylla – the suggestion that there are or might be absolutely un-settleable moral disputes. Scylla, the minimalist discovered, is inconsistent with the first two minimalist claims – namely that moral discourse is truth-apt and that moral truth is epistemically constrained. Hence – according to the minimalist – there is never a point where we can just throw up our hands and say: ‘Well, there is no objective truth of the matter here. We will have to find our own truth’. On the contrary, according to the minimalist, it is part of the claim that moral discourse is truth-apt that we must always continue to seek an objective resolution to actual disputes. Or, to put this using Price’s terminology: It makes no sense to suggest that there are or could be evaporative *inferential* disagreements – that is, cases where neither disputant is mistaken about apparently contradicting linking-principles. Price’s form of relativism (parochialism) is inconsistent with minimalism.

But, what about the second central claim of the minimalist. Namely that moral discourse does not exhibit Cognitive Command – that we ought to remain agnostic about whether it is a priori that moral disagreements involve cognitive shortcomings. Or to use again Price’s terminology, the minimalist thesis about truth without Cognitive Command is the thesis that *we must remain agnostic about whether it is a priori that* there are no (and could be no) evaporative inferential disagreements. One way of seeing this minimalist claim is as an attempt to capture the underlying truth in relativism. In particular, the reason for the minimalist’s agnosticism is exactly the fact that a subtle form of relativity appears to play a role in moral discourse.

That is, the affective grounding of moral discourse means that the affective natures of the subjects play an important role in moral epistemology. And the subjective nature of a subject is in turn influenced by the culture in which he is immersed (among many other influences) and, hence, relative to such influences. Therefore, the epistemology of moral discourse – at least given what we have so far

discovered – cannot be such that we can have an *a priori* guarantee that any disagreement is cognitively blameless, since, crucially, the disagreement may ultimately be down to non-cognitive (e.g. cultural) differences in the affective natures of the subjects.

Hence, minimalism can – as can be seen from this way of paraphrasing the position – appropriately be called a form of relativism. Therefore, to sum up, minimalism is a form of relativism that does not sanction relativistic practice. On the minimalist's account it always makes sense to continue to *seek* consensus over moral matters even though we cannot be sure *a priori* that we will get it. Likewise, the minimalist's form of relativism does not have problems with explaining the categoricity of moral demands. This gives his form of relativism a dialectical advantage over the strong – parochialist – form of relativism.

Chapter Six

Minimalism, Supervenience and Naturalism.

So far, the argument of this thesis has focussed on making sense of a cognitivist, *sui generis*, but non-objectivist, account of moral discourse – that is, making sense of moral truth without Cognitive Command. However, even if sense can be made of this option, it might appear that there is still the possibility that the case for moral minimalism – failure of moral Cognitive Command in particular – may be undercut by the establishment of some sort of necessary link between moral discourse and objective naturalistic discourse either through the defence of a sufficiently strong form of substantive moral-on-natural supervenience¹⁸⁷ or through a successful case for a substantive moral naturalism. In this chapter, I investigate whether such necessary links, if made good, would threaten to impose satisfaction of the Cognitive Command constraint on moral discourse.

To aid my investigation, I shall avail myself of and develop a notion of *disputational supervenience* originally defined by Wright.¹⁸⁸ A discourse – to give a preliminary definition – disputationally supervenes on another discourse if and only if disagreements in the supervening discourse are rationally intelligible only if there is a further disagreement in the supervened-on discourse. The leading question in the following will be whether substantive moral-on-natural supervenience or the truth of some form of substantive moral naturalism establishes this relationship between moral and natural discourse. If one of these phenomena succeeds in doing this, it would seem that an a priori guarantee of there being a cognitive shortcoming on at least one side of every basic moral dispute is in the offing. That is, since the underlying naturalistic dispute is guaranteed a priori to involve such a cognitive shortcoming – I am assuming here that the relevant naturalistic discourses exhibit Cognitive Command – the disputationally supervening moral dispute is also guaranteed a priori to involve a cognitive shortcoming.

¹⁸⁷ By 'substantive moral-on-natural supervenience' I shall understand the supervenience-relation that holds between moral *properties* and natural *properties*. There is universal agreement that some sort of necessary relation holds, but as we shall see in the following there are disagreements about the actual strength of this relation.

¹⁸⁸ See Wright (1992), p. 155-56.

To anticipate, I shall argue that neither substantive moral-on-natural supervenience nor any form of moral naturalism presents any immediate threat to moral minimalism. I shall begin the chapter by refining the notion of disputational supervenience for the purposes of the discussion in the rest of the chapter (section 1). The rest of the chapter will then be divided into two discussions. First, there is the question of whether the phenomenon of substantive moral-on-natural supervenience enforces moral-on-natural disputational supervenience. I shall argue that *actual* substantive moral-on-natural supervenience is of a very weak ‘intra-world’ kind – as argued by Hare and Blackburn. A further question that is addressed – deriving from Blackburn's supervenience-argument against the moral Realist – is how the phenomenon of substantial supervenience can be *explained* by the minimalist platonist (sections 2, 3 and 4). I also dismiss suggestions that actual substantive moral-on-natural supervenience is of a stronger, inter-world, kind. I argue that, anyway, none of the stronger principles threaten to impose moral-on-natural *disputational* supervenience (sections 5 and 6).

In the second discussion of this chapter, the leading question is whether some form of moral naturalism – reductionism broadly defined – may somehow bestow objectivity on morals via enforcing moral-on-natural disputational supervenience. I, first, attempt to get the issue of the relation between moral-on-natural disputational supervenience and moral naturalism into sharp focus and to provide some taxonomy of reductive naturalisms (section 7). I then turn to the prospects of analytical naturalism imposing moral-on-natural disputational supervenience. I, first, discuss the attempt to disqualify all analytical naturalistic accounts with the open question argument (section 8). I then argue that the forms of analytical naturalism that survive the open question argument – I take Frank Jackson's moral functionalism as an example – cannot for several reasons impose moral-on-natural disputational supervenience (section 9). Finally, I turn to arguing that there are even more reasons for doubting that ontological – i.e. non-analytical – naturalism can save moral discourse from non-objectivism via establishing the disputational supervenience of moral discourse on naturalistic discourse (section 10).

1. *Disputational supervenience and Cognitive Command.*

The thought that threatens moral minimalism is that if moral discourse is linked in a sufficiently strong way to naturalistic discourse, then the objectivity of the naturalistic discourse – which arguably *does* satisfy the Cognitive Command constraint – will somehow 'rub off on' moral discourse. As already mentioned, this will turn on whether the link is strong enough to amount to disputational supervenience. Let me, first, define this notion of disputational supervenience more precisely. *Disputational supervenience* as defined by Wright is this:

A discourse, D, disputationally supervenes on discourse, B, if and only if it is a priori that a D-dispute is rationally intelligible only if there is a further difference of opinion formulable within B.

This contrasts with *disputational purity* – the feature a discourse has exactly when there need be *no* further dispute formulable in any other discourse in order to rationally sustain a basic dispute within the discourse.

So, the idea is that some discourses have and some discourses lack the following feature: That there is a conceptual guarantee that if one is involved in a basic disagreement within the discourse – a dispute over linking-principles – and if one is rational, then one is also involved in a further dispute over some issue formulated in the terms of another discourse. This feature is what we are looking for in moral discourse, if what we are after is an argument that there is an a priori guarantee that moral disagreements involve a cognitive shortcoming on the part of one or both disputants – i.e. an argument for Cognitive Command in morals – based on the existence of some sort of necessary link between naturalistic and moral discourse.¹⁸⁹ To see how this feature of disputational supervenience may result in moral discourse inheriting the degree of objectivity – understood here as the satisfaction of the Cognitive Command constraint – that naturalistic discourse is presumed at least to have, consider the following argument.

Given moral-on-natural disputational supervenience, it will be a priori knowable that any moral dispute will involve some further difference of opinion

¹⁸⁹ Of course, this is not the only way of establishing moral Cognitive Command. Morals might exhibit Cognitive Command in its own right and not via necessary links to naturalistic discourse.

formulable in naturalistic terms. And with regard to this further dispute, we have an *a priori* guarantee that *it* must involve a cognitive shortcoming on the part of one or both disputants, since we are presuming that naturalistic discourse exhibits Cognitive Command. Thus, we can come to know *a priori* that whenever there is a dispute formulable in moral discourse, there is a further difference of opinion formulable in naturalistic discourse that must involve a cognitive shortcoming. Therefore, disputational supervenience on a discourse that satisfies the Cognitive Command constraint guarantees the satisfaction of the Cognitive Command constraint for the disputationally supervening discourse.

Now, the definition of disputational supervenience above is crucially vague in that the notion of rational intelligibility comes in different strengths. Exactly how much is required of a moral disputant in order to be rationally intelligible? Ought he to consciously believe everything that follows logically from what he consciously believes – or ought he even consciously believe every conceptual truth that he implicitly accepts given that he masters the concepts that he masters? To get a more precise grip on the notion of disputational supervenience, I want, first of all, rephrase the definition of it in terms of the notion of a *rational commitment*. Thus:

A discourse, D, disputationally supervenes on discourse, B, if and only if it is *a priori* that a D-dispute rationally commits the disputants to a further difference of opinion formulable within B.

Then the question becomes: What are we rationally committed to as competent practitioners in moral discourse? We can here choose to work with a *lax* and a *strict* notion of rationality.

If what is meant is merely lax rationality, then we will not require of someone with a mastery of a set of concepts that he believes *everything* that follows logically or conceptually from what he believes while applying these concepts, but only what is *relatively obvious (logical and conceptual) consequences* of what he believes. (For example, a rational practitioner with the arithmetical concept of plus will be rationally required to believe that $5+7=12$, but not – at least not without pen and paper or a pause for thought – that $1139960+12005976 = 13145936$). This is, arguably, the way we normally use the notion of rationality. If this is the notion that we choose as relevant to our definition of disputational supervenience, then it is doubtful that moral

discourse will disputationally supervene on any other discourse. It would seem that there are examples of radical moral disputes where it is certainly not obvious that there is an underlying naturalistic dispute present. At least, it would take hard philosophical thinking to realise that there was, if there was any.

We may, however, choose to work with a stricter notion of rationality in our definition such that rational disputants are required to believe even *un-obvious* logical and conceptual consequences of the beliefs they have and the concepts they possess. If this is our choice for explicating the notion of rational commitment, the prospects for moral-on-natural disputational supervenience are clearly brighter. In particular, there are three ways in which un-obvious conceptual truths about moral discourse might help to establish disputational supervenience – three candidates for being the feature of moral discourse that ensures that it is not rationally intelligible to have a moral dispute without having a naturalistic dispute. I shall later argue that none of the candidates succeed, but let me, first of all, list them:

(1) Perhaps something can be made of the relatively un-obvious conceptual truth that moral discourse substantively supervenes on naturalistic discourse.¹⁹⁰ I shall discuss three different attempts to establish moral-on-natural disputational supervenience working with substantive supervenience-principles of varying strength. (2) There is possibly some very complicated – hence un-obvious – way of *analysing* moral terms in naturalistic terms. That is, it might be a priori and necessarily, but un-obviously, true that moral concepts are equivalent to naturalistic concepts – as argued by what I shall call the analytical naturalist. (3) Or, finally, it may be that it is un-obviously a priori knowable that moral terms function rigidly – like natural kind terms. Hence it will be un-obviously a priori knowable that *some* underlying natural property is (metaphysically) necessarily identical with a given moral property (where the exact

¹⁹⁰ Thus, a substantive moral-on-natural supervenience claim (whatever its form) is generally held to be a *conceptual* truth – a moral platitude, an a priori and necessary truth. For example, someone who regards it as possible that two actions are similar in all relevant descriptive respects, yet holds, for example, that the one action is morally right and the other is wrong, will simply be thought of as someone who does not master the concept of rightness. This sets out substantive moral-on-natural supervenience from other supervenience-phenomena. The water-property might be held to supervene on the H₂O-property, for example, but this is surely something that we can only know a posteriori. For doubts – to which I am sympathetic – about whether we need to invoke metaphysical necessity to explain the relation between our concepts of water and H₂O, see Hare (1984), pp. 12-15.

identification of the underlying natural property is an a posteriori matter). This is the strategy of what I shall call the ontological naturalist.¹⁹¹

Which notion of rationality – lax or strict – should we work with in our definition of disputational supervenience? Well, as already remarked, the lax notion is certainly the one that is most in line with the way we actually would use the term ‘rational commitment’. To choose the strict notion would be to require of a competent user of moral concepts that he also possessed quite sophisticated reflective – philosophical – knowledge of conceptual truths and that is surely too strong a demand to place on ordinary moral practitioners. So, in a way our investigation of the threat from moral-on-natural disputational supervenience to minimalism could stop right here, since the lax notion of rational commitment is very unlikely to deliver moral-on-natural disputational supervenience.

However, I want to give the would-be moral objectivist some more rope by working with the strict notion of rationality in the following for two reasons. First, I want to show that even with this concession the objectivist is not going to succeed in establishing that moral-on-natural disputational supervenience holds. Secondly, it is worthwhile to show that moral-on-natural disputational supervenience is not guaranteed even when the disputants are philosophically sophisticated, since we may imagine a moral discourse that has been enriched by philosophical progress. In other words, it would still be an interesting question whether a philosophically enlightened moral practice could exhibit Cognitive Command. So, I am – reluctantly – conceding to the would-be objectivist that we, in the following, should work with the strict notion of rationality in our definition of disputational supervenience. That is, participants in moral discourse are assumed to be aware of all conceptual truths – even the relatively un-obvious ones – that are implicit in practice with moral concepts.

The matter in focus in the following is therefore whether, with disputational supervenience defined in this way, actual substantive moral-on-natural supervenience or some form of substantive moral naturalism is going to be able to secure this feature for moral discourse.

¹⁹¹ See section 7 for more detailed definitions of analytical and ontological naturalism.

2. *Actual substantive moral-on-natural supervenience.*

In the following five sections, I address the question whether actual substantive moral-on-natural supervenience is strong enough to enforce the corresponding disputational supervenience or in other ways to cause problems for the minimalist. There are many types of supervenience-relation and there is no consensus, as will become evident in the following, about which applies in the moral case. In this section and section 4, I take on the question of what principle of supervenience is implicit in actual moral practice and I argue that it is of a weak intra-world kind – or at most of a very weak, counterfactual, kind. And I show that this principle does not impose moral-on-natural *disputational* supervenience. In sections 3 and 4, I also present, and attempt to *meet* on behalf of the minimalist, the explanatory challenge to moral Realists – or moral Platonists – famously posed by Blackburn: How can the minimalist *explain* the supervenience-phenomenon – and, in particular, this without invoking some form of substantive naturalism? Finally, I consider and dismiss two alternative proposals made by McFetridge and Jackson for what actual substantive moral-on-natural supervenience might be – a weak and a strong inter-world principle. And I argue that none of these succeed in imposing moral-on-natural *disputational* supervenience anyway (sections 5 and 6).

Though there is disagreement about which substantive moral-on-natural supervenience principle is true, it is safe to say that most philosophers accept what I – given that I agree with these philosophers – shall simply call *actual substantive moral-on-natural supervenience* or just (S). This is an *intra-world*, cross-object supervenience principle. If we take M to ascribe moral properties and N to ascribe natural properties, then the intra-world, cross-object substantive moral-on-natural supervenience thesis states that if, in some possible world, two objects share all natural properties, then, *in that possible world*, they share all moral properties. Accordingly, the following principle is accepted as a conceptual truth:

- (S) For any objects, x and y, and for any possible world, w: if x and y share all naturalistic properties in w then x and y share all moral properties in w.

Or, in McFetridge's useful formalisation and tabulation (where Nx is a naturalistic predicate and Mx is a moral predicate):

$$(S) = (XYWW) (\forall w)(\forall x)(\forall y) [(\forall N) (N_{xw} \leftrightarrow N_{yw}) \rightarrow (\forall M) (M_{xw} \leftrightarrow M_{yw})]$$

Thus, what the substantive supervenience principle (S) says is thus just that one must not be inconsistent in one's ascriptions of moral properties in the sense that naturally wholly similar cases must be treated morally similarly.

Now, actual substantive moral-on-natural supervenience – in the shape of (S) – is too weak to enforce disputational supervenience. To see this, consider that there is nothing wrong in assuming that the following dispute is rationally intelligible. Imagine that A believes that Mx and N^*x and B believes that $\text{not-}Mx$ and N^*x (where M is a moral predicate and N^* is a maximal naturalistic predicate – that is a full naturalistic description to which nothing *can* be added).¹⁹² Assume also – in line with the above – that both A and B are aware of the un-obvious conceptual truth of (S). In other words, imagine that A and B are involved in a moral dispute without being involved a corresponding naturalistic dispute – i.e. imagine a counterexample to the moral-on-natural disputational supervenience thesis. Now, there is no problem in assuming that A and B in this scenario are both rationally intelligible. They can rationally comply with (S) while hanging on to their respective beliefs. All that is rationally required of A – given knowledge of (S) – is that he does not claim that some other object, y , does not satisfy M while satisfying N^* . Similarly, B is required not to claim that some other object, y , has N^* and has M too. Thus, in no way are A and B rationally required to disagree over whether x has N^* .

So, actual – intra-world – substantive supervenience is not going to do the job of establishing that moral discourse disputationally supervenes on naturalistic discourse. But, there are many other types of supervenience on offer – in particular *inter-world* supervenience theses.

3. *Explaining substantive moral-on-natural supervenience.*

Before turning to examples of inter-world supervenience theses, however, let me digress briefly to consider Blackburn's suggestion that (S) in a different way *does* pose a problem for any moral Platonist. I want to make sure that a version of his argument

does not apply to the minimalist platonist. Blackburn argues that (S) coupled with the Moorean insight that there are no entailments from natural descriptions to moral descriptions – the insight allegedly sparked by the discovery of the naturalistic fallacy – spells trouble for the moral realist.¹⁹³ As I shall argue, the argument is ultimately successful in that it eliminates the Platonist, and, adding a little to Blackburn's original argument, the ontological naturalist realist. Blackburn advances the following two theses as conceptual truths:¹⁹⁴

(S) For any objects, x and y , and for any possible world, w : if x and y share all naturalistic properties in w then x and y share all moral properties in w – i.e. (XYWW) above:

$$(\forall w)(\forall x)(\forall y) [(\forall N) (N_xw \leftrightarrow N_yw) \rightarrow (\forall M) (M_xw \leftrightarrow M_yw)]$$

(L) There is an ascription of a moral property to an object that is true and that is such that no ascriptions of natural properties of this object strictly imply¹⁹⁵ the ascription of the moral property to the object, i.e.:

$$(\exists x)(\exists M) [(M_xa \ \& \ \neg(\exists N) (N_xa \ \& \ (\forall w) (N_xw \rightarrow M_xw))]$$

(L) could have been the claim that *no* moral facts are strictly implied by natural facts, but this stronger principle is not needed in what follows. As it shall become evident later, according to Blackburn, the principle (L) distils Moore's argument against moral naturalism.

Now, according to Blackburn moral realism is incompatible with these two (alleged) conceptual truths. To see why he can claim this, observe, first, that we may define a naturalistic predicate N^* as the predicate such that ascribing N^* truly to an object is to give it a fully determinate naturalistic description – a description so fully determinate, in fact, that no further natural description can be truly added to it. The

¹⁹² I say more about maximal naturalistic properties in the next section.

¹⁹³ Blackburn (1984), p. 182-87.

¹⁹⁴ For the former, see Blackburn (1993), p. 131ff and for the latter p. 116 and 120.

¹⁹⁵ Following McFetridge (1985), p. 251, I define 'strict implication' as follows: p strictly implies q if and only if necessarily [p implies q]. The necessity involved here may be both conceptual/logical and metaphysical. On the other hand, I define 'entailment' as follows: p entails q if and only if it is *conceptually/logically* necessary that [p implies q]. Notice that the latter definition of the word 'entailment' is different from, for instance, Jackson's. Jackson uses the word 'entailment' such as to include derivations based on a posteriori necessities. Jackson (1998), p. 25. Though he does also recognise that there is a use – similar to the one I make – of the word 'entails' that can be paraphrased as 'conceptually entails' or 'a priori entails'.

claim is then that all objects in the actual world – describable in naturalistic terms – have such a 'maximal naturalistic property':

$$(NMA) \quad (\forall x)(\exists N) [N_xa \ \& \ (\forall w)(N_{xw} \rightarrow (\forall N') (N'_{xa} \leftrightarrow N'_{xw}))]$$

In other words, every object has a naturalistic property N^* such that if the object has this property in the actual world as well as in another possible world, then the object in the actual world will share *all* its naturalistic properties with the object in this other possible world.¹⁹⁶

Now, (S) rules out possible worlds in which, say, object x has naturalistic property, N^* , and moral property, M , and object y has naturalistic property N^* , but does not have moral property M . That is, (S) bans mixed possible worlds – to use Blackburn's phrase. But, then Blackburn asks: How can the realist account for this ban? The Platonist version of the realist appears not to be able to do this. He believes that moral properties are real and autonomous – existing in a realm that 'floats free' of nature. But, why should it not then be possible – on the Platonist's account – that moral properties were distributed unevenly among objects with the property N^* ? Here the realist might try to go naturalistic – while of course loosing his right to the title 'Platonist' – and argue that there is a necessary connection between having property N^* and having property M . But, according to Blackburn, given that Moore has convinced us that we must accept the truth of (L), we must forever give up on this substantive naturalistic road. Thus the only option for a realist is to be a Platonist and allow for possible worlds in which *no* object that can be fully described naturalistically by N^* has moral property M . So, why not also – on the Platonistic/realistic account – allow for possible worlds in which *some* objects have N^* and M and others have N^* and not M , i.e. allow for mixed possible worlds? The Platonist (and perhaps every other non-naturalist realist) appears to leave it *utterly mysterious* why there is a ban on mixed possible worlds as the one imposed by (S).

However, to see why this argument is not immediately successful against *all* forms moral realism (or more precisely, is not successful against all forms of moral

¹⁹⁶ I use McFetridge's formalisation and name. See McFetridge (1985), p. 250. That (NMA) holds is important to keep in mind, since this excludes, for instance, the possibility that x has all the naturalistic properties that y has, yet differ in moral status simply because y has some *further* natural

realism that are not substantive analytical naturalisms), consider what might convince us that (L) was true.¹⁹⁷ Blackburn attempts to back up (L) by appealing to the following principle:¹⁹⁸

(E') There are some moral propositions that are true, but whose truth is not entailed by any naturalistic facts about their subject.

Blackburn's acceptance of (E'), it seems, is in turn just his acceptance of the Moorean claim that it is impossible to give a naturalistic *analysis* of moral predicates. Crucially, however, (E') speaks of there being no *entailment* from natural to moral facts where (L) spoke of there being no *strict implication*. (As already mentioned, entailment is here defined as either logical or conceptual deducibility, that is, a necessary implication that is in addition knowable a priori). So, (L) is backed up by (E') – and *eo ipso* Moore's arguments – only if we can accept generally that where there is no entailment, there is no strict implication. But, this is, of course, a controversial claim post Kripke's and Putnam's 'invention' of a posteriori necessities.

In other words, Blackburn's argument appears not to be effective against realists who are willing to countenance merely a posteriori knowable necessities linking moral and natural facts. That is, Blackburn did not – at least in his original formulation of his supervenience-argument against the realist – take into account the possibility of being a *non-analytical* moral naturalist. That is why McFetridge is right to claim that Blackburn's argument is only partly successful against the moral realist. It is successful against the moral Platonist who rejects any necessitation claims linking moral and natural properties. But, it is not fatal to the realist who is also a non-analytical – what I shall call an ontological – naturalist, as for instance a Cornell realist.

However, McFetridge does not succeed ultimately in defending the moral realist against Blackburn's supervenience-argument. I now attempt to criticise McFetridge by arguing that Blackburn *can* extend his argument from substantive moral-on-natural supervenience – in the shape of (S) – against moral realism so that it

properties. (NMA) ensures that one of the naturalistic predicates applying to an object will be a complete description of it such that no further true description *can* be added.

¹⁹⁷ The following diagnosis of Blackburn's argument is made by McFetridge (1985).

¹⁹⁸ Blackburn (1993), p. 120.

also becomes effective against the ontological naturalist. We must grant that the ontological naturalist is able to explain the ban on mixed possible worlds by appealing to a posteriori necessitation-claims. But, only to land himself in trouble by no longer being able to explain the *a priori* of (S). Thus, the ontological naturalist cannot make sense of a priori supervenience by appealing to a posteriori links between properties.¹⁹⁹ So, the ontological naturalist ends up with a difficult explanatory problem – different from the one facing the Platonist, but devastating none the less.

So, it would seem that Blackburn's argument with an added point, pace McFetridge, is efficient against *all* forms of moral realism. We must therefore ask whether it also has consequences for the minimalist platonist.

4. Moral minimalism and substantive moral-on-natural supervenience.

There are in fact *two* questions that have been raised in the foregoing to which the minimalist owes an answer. In this section, I attempt to give answers on behalf of him. Firstly, there is the question that has just been raised of whether the minimalist can *explain* the truth of the supervenience-thesis, so as not to fall prey to Blackburn's – ban on mixed possible worlds – argument against the moral realist. But, secondly, there is the independently interesting question about what *type* of substantive supervenience principle has the status of a conceptual truth in *actual* moral practice. Later in this section, I shall argue that we – including of course the minimalists among us – ought to regard (S) as being this principle.

What is the minimalist response to Blackburn's argument from mysterious bans on mixed possible worlds? Blackburn answers on behalf of his own quasi-realist essentially by making a distinction between seeing the substantive moral supervenience thesis as a *metaphysical* principle and seeing it as just a useful feature of the way we moralise. The quasi-realist holds that supervenience is a result of the function moralising has in our lives. It has become part of being a competent moral judge that one respects supervenience as a constraint on competent ascriptions – or

¹⁹⁹ Blackburn recently (see his (1998), p. 316) hinted at this problem for the ontological naturalist. See also his (1993), the addendum p. 145-48 for Blackburn's most recent comments on supervenience.

rather 'projections' – of moral properties because it serves the overall purpose of moralising.²⁰⁰

But, I will argue, the minimalist can give a similar answer. He can hold that (S) is just a product of the discipline that governs moral discourse. It is just a (core) platitude that treatment of wholly naturalistically similar cases in morally different ways is prohibited – a norm that any competent user of moral concepts respects. This answer appears not to satisfy Blackburn, however (assuming that he is here using 'minimalism' in a way that includes our non-deflationist form of minimalism):

'In the case of ethics, supervenience seems to be built into the discourse – it is analytic – and I still see quasi-realism as the only genuine explanation of this. 'Minimalist' positions, for example, simply have to take it for granted, and this, to me, is letting the spade be turned far too soon'.²⁰¹

I think we should grant Blackburn that *just* to observe that (S) is platitudinous is no explanation. Thus, I do not think Jackson does enough to explain his a priori acceptance of the strong supervenience-principle, he accepts, (i.e. the principle (XYWW') that I shall discuss later in this chapter) with no more than the following words:

'It is an implicit part (if it were explicit, the matter would not be philosophically controversial) of our understanding of ethical terms and sentences that they serve to mark distinctions among the descriptive way things are'.²⁰²

Of course, what is needed is an explanation of *why* this has become an implicit part in our moral practice. However, I do not see this as an insuperable challenge to moral minimalism. In particular, what I argued in chapter five – that minimalism can be complemented by an explanatory naturalistic theory – will enable the minimalist to explain substantive moral-on-natural supervenience and help to determine which principle is the one that is actually implicit in our moral practice. Let me therefore turn to the explanation on behalf of the minimalist of why the supervenience principle is a conceptual truth.

As Hare points out, the substantive moral-on-natural supervenience constraint is part of – entailed by – the general universalisability constraint for moral claims.²⁰³

²⁰⁰ Blackburn (1984), p. 186.

²⁰¹ Blackburn (1998), p. 315.

²⁰² Jackson (1998), p. 125.

Universalisability, thus, comprises two constraints on responsible moralising: First of all, the following is required of any competent user of a moral term. If he makes a moral particular – that is, occasion-specific – claim of some kind, then he ought on all future occasions to comply with *a* principle that backs up this particular claim. Moral judgement must in this way be principled. More precisely, he has to comply with a universal moral statement which, in conjunction with a naturalistic premise, entails the particular verdict. This is what sets out moral prescriptions – moral oughts – from mere imperatives according to Hare. But, this trivially means that the following principle holds true:

Necessarily, if *a* has moral property *M*, then there is a universal principle:
 $(\forall x) (Nx \rightarrow Mx)$ and a particular naturalistic truth, *N_a*, such that *M_a* is entailed by these.

In other words, there is some universal rule governing the use of every moral predicate such that if it applies to *a* then it applies to *b* in so far as *a* and *b* are naturalistically qualitatively identical – in particular, in so far as they are identical with respect to the properties ascribed by *N* in the antecedent of the universal principle. But, this is, of course, just another way of saying that moral properties supervene on natural properties in the sense of supervenience given by (S).

The second part of the universalisability constraint is that the universal principle backing a particular moral claim must not involve any reference to individuals (in general, particulars or spatio-temporal locations). I cannot intelligibly offer as justification for a moral claim some ad hoc reason that specifies, for instance, that a certain moral predicate applies to all individuals of a certain naturalistic description except me.

But, if substantive moral-on-natural supervenience is simply a consequence of the feature of moral practice that we are rationally committed to comply with some moral principle – an impersonal principle – whenever we make a moral particular claim, then there is no mystery about the phenomenon for the minimalist. Let me for a short while re-enter into the mode of the explanatory naturalist to see how the universalisability constraint fits the account of moral practice given by him.

²⁰³ See for instance Hare (1984), p. 3 and 9, Hare (1963), chapter 3 and Hare (1981), chapter 6. The following account of supervenience is very much inspired by his.

The challenge is to understand how the demand for universalisability is important for achieving what we achieve by engaging in moral discourse. The explanation could go as follows. The feature of moral discourse that there is a call for a principled backing of moral particular claims is exactly the feature that *creates the space for moral debate*. According to this requirement of principle, if two people disagree about some particular moral judgement – e.g. that *that* act is wrong – then they are both required to refer to universal statements in support for their respective claims. But, then they are both vulnerable to further criticism. A can attempt to produce counter-examples to B's moral principle and vice versa. Both disputants must then try to remain consistent, to give a coherent story, about the principles (of varying generality) they hold true and the moral verdicts about particular cases they give. This can result in revisions in both principles and particular verdicts by both participants to the debate. And the possibility, but not guarantee, of convergence and change – adjustment and perhaps improvement – in moral beliefs is created. So, we see how the demand for universal reasons in the context of moral debate – that can be seen as the rationale behind the supervenience and universalisability requirements – can be explained by reference to the overall function of moralising.²⁰⁴

Likewise, it is easy to see why the demand for *impersonal* reasons is a necessary part of our use of moral concepts. Ad hoc principles that explicitly serve the interests of particular individuals or groups are not likely candidates for principles that can serve as the basis for a consensus. Thus, if the purpose of moral discourse and truth is partly to generate a drive towards a moral consensus – a moral code that can co-ordinate life in a community generally – then moral principles explicitly tailored to fit the interests of particular agents must be regarded as in-admissible for the purposes of moral justification. Remember that co-ordination achieved through normative debate, as opposed to through dogma or authority, is co-ordination based on individuals in a community accepting a common set of moral principles. But, such a general acceptance is unlikely to occur if the participants perceive that they are not going to benefit from the consensus – if the principles being discussed in the debate are biased, tailored to suit the interests of particular individuals. (I shall later in this

²⁰⁴ I have, of course, said much more about the function or purpose of moral debate from an explanatory naturalistic perspective in chapter five.

chapter say more from the explanatory naturalist perspective about the substantive supervenience-phenomenon).

So, the minimalist *can* explain the supervenience-phenomenon – in answer to Blackburn's challenge – once it is granted that he can complement his theory with methodologically naturalistic explanations of features of moral practice. But, this still leaves the second question posed above. That is, the one about *which* principle of substantive supervenience is true – which principle has platitudinous status within actual moral practice. I already implicitly assumed in the account just given that the supervenience-principle that has platitudinous status in actual moral practice is (S). That is, what gives us supervenience is the demand for moral justification by universal principles and there is nothing in the phenomenon of reason-giving – and in the way having to give universal reasons generates the opportunity for normative debate – that calls for these principles to be claimed to be *necessarily* true. A stronger supervenience principle could, for instance, be defended by providing an argument for the following principle (a principle equivalent to McFetridge's supervenience-principle (XXWA) below that I shall discuss separately later in this chapter):

Necessarily, if *a* has moral property *M*, then there is a *necessarily* true universal principle: $\Box(\forall x) (Nx \rightarrow Mx)$, and a particular naturalistic truth, Na , such that Ma is entailed by these.

Thus, on this proposal, the type of justification called for in moral discourse would be justification derived from postulated *necessary* universal principles. But, the question is evidently why this should be true. Why should we not, when adopting justificatory moral principles, be willing to accept that these principles *might* not have been true? All that is needed, it would seem, is that the principle holds true in *this* world (and perhaps nearby possible worlds, as I shall shortly argue). For the purposes of generating reasoned moral debate, there is no need to back up particular verdicts with claims about *necessarily* true moral principles. But, then there is no case to be made for stronger supervenience principles (like McFetridge's (XXWA) or Jackson's (XYWW') that I shall discuss later) from an explanatory perspective.

Let me defend this claim in some more detail by arguing explicitly against McFetridge's claim that the stronger principle (XXWA) is the principle that is a platitude. I.e.:

$$(XXWA) \quad (\forall w)(\forall x)[(\forall N) (Nxa \leftrightarrow Nxw) \rightarrow (\forall M) (Mxa \leftrightarrow Mxw)]$$

Since Jackson's stronger supervenience-claim entails McFetridge's claim, my criticism will also apply to him. McFetridge claims that it is 'almost self-evident' that (XXWA) is the principle inherent in common moral practice.²⁰⁵ I shall advance two arguments against accepting McFetridge's principle as a formulation of actual substantive moral-on-natural supervenience. First of all, I argue that we ought to adopt a *cross-object* supervenience principle. Secondly, if we are to adopt an inter-world principle, we ought only to be interested in a *restricted* range of nearby possible worlds.

So, first, why ought we insist that actual moral-on-natural supervenience is a cross-object principle? The minimalist can, of course, against McFetridge's claim that (XXWA) is 'almost self-evident', just insist that *his* intuitions point towards an intra-world, cross-object, supervenience principle. But, let us test our intuitions before drawing any conclusions. McFetridge tries to call up intuitions favouring his preferred version of the supervenience-principle with a quote from Hare:

'If I said 'Smith acted rightly in giving her the money, but he might have given her the money, and in all other respects acted similarly, except that his act was not right', I should invite the comment: 'But, how could the rightness of the act disappear like this? If the act, motives, circumstances, &c., were all the same, you would be bound, logically, to judge it right in the hypothetical case as you did in the actual case. The actual action couldn't have been right and the hypothetical situation not right, unless there had been *some* other difference between the action, or their circumstances, or their motives, or something else.'²⁰⁶

Hare's point here rings true to me. The question is whether McFetridge's interpretation of the intuition is the right one when he sees it as establishing the truth of (XXWA). We might here immediately ask why Hare – given his insistence on *impersonal* reasons – could not have changed his example slightly and shown that it is equally odd to suggest that *Jones* – that is, another person – who, say, also gave her the money in exactly the same circumstances, with the same motives etc., did not act rightly. In that case, Hare's intuition with the appropriate addition in fact supports a *cross-object* principle – an XY-principle – like (XYWA), rather than (XXWA). And, reading Hare this way is also supported well by other parts of Hare's writings where

²⁰⁵ McFetridge (1985), p. 249.

he focuses on the importance of role-reversal-test.²⁰⁷ In other words, if McFetridge wants to use Hare's intuition to back up his account of supervenience, he ought to change his position and argue for a cross-object supervenience principle.

Now to the second point against McFetridge. I shall try to challenge his conclusion in two further ways relating to the claim that we ought to adopt an inter-world principle. One by attempting to challenge Hare's intuition and one by challenging McFetridge's interpretation of it. Only the latter challenge will prove successful.

We might, first of all, attempt to resist McFetridge's principle by challenging Hare's intuition. More precisely, we might advocate a strong form of what might be called *moral actualism* – the idea that only what happens in the actual world has moral significance. According to this view, there is nothing worrying about Hare's maverick moralist. It does not matter, that he is willing to deem acts right in the actual world, yet not deem naturally similar acts right in other possible worlds. We do not (or at least ought not) according to this view care about what befalls our counterparts, or the counterparts of our fellow humans, in other possible worlds. What worries us ought always to be what will happen to ourselves and our fellow humans in *the actual* world. Therefore, what should agitate us is what Hare's moralist will judge about actual cases. The argument would be that the purposes of moral debate are served perfectly, if we restrict our attention to the people that we *actually* affect with our actions and whose actions *actually* affect us.²⁰⁸

However, this suggestion – to become moral actualists – does not quite eliminate the feeling that Hare's maverick moralist is somehow misusing the term 'right' in the quote above. It seems, to me at least, that we *are* normally interested in the other disputant's being willing to apply the principles he advocates in hypothetical situations – even given that we do not care about our counterparts in other possible worlds. Thus, the minimalist cannot escape Hare's intuition and must provide an explanation.

So, the question instead becomes: Is McFetridge interpreting the intuition correctly when he sees it as establishing the truth of an inter-world principle like

²⁰⁶ Hare (1952), p. 153 and McFetridge (1985), p. 249.

²⁰⁷ The next quote from Hare in this section illustrates this point well.

(XXWA)? It certainly seems that his gloss on it differs from the intentions of Hare himself. Hare appears to have a special type of non-actual worlds in mind when he speaks of hypothetical cases:

'It is always legitimate, in order to apply to moral argument the requirement of universalizability, to imagine hypothetical cases which really are, apart from the fact that the roles of the people concerned are reversed, precisely similar in the relevant respects to the actual case being considered; and this may properly be done, however fantastic the assumptions that have to be made, in matters that do not affect the moral issue, in order to make the hypothetical case seem possible'.²⁰⁹

That is, Hare is – in adherence to the universalizability-requirement – only interested in the nearby possible worlds in which the roles of the people involved in some situation to be judged morally are reversed.²¹⁰ But, this falls far short of wanting to consider *all* possible worlds – as McFetridge interprets him.²¹¹ Hare has something weaker than necessity in mind and talks instead about nomological principles and subjunctive conditionals and draws parallels to causal laws:

'However, it does seem that, at least in the moral and the causal cases, the universal premise, though it does not need to hold necessarily, has to be nomological in character (as the premise about cyanide was: it implies that if any food had contained it, it would have been poisonous). I have argued that the universality of moral principles requires them to apply to hypothetical cases as well as actual, and the same seems to be true in the causal case. It is easy to confuse nomologicality with holding necessarily, but they must be distinguished'.²¹²

So, according to Hare – and I should add, quite plausibly – the universal principles that can and should be used to justify moral claims are more like subjunctive conditionals. Hare's comparison with causal laws is instructive. When I claim that it is a natural law that cyanide, if ingested, causes poisoning (in normal human beings), I do not mean to say that it is true in all possible worlds that cyanide, if ingested, causes

²⁰⁸ The argument just presented may, thus, only justify the following supervenience-principle – even weaker than (S): $(XYAA) (\forall x)(\forall y)[(\forall N) (Nxa \leftrightarrow Nya) \rightarrow (\forall M) (Mxa \leftrightarrow Mya)]$

²⁰⁹ Hare (1963), p. 44. See also Hare (1981), p. 114.

²¹⁰ Which incidentally also indicates that he is after a cross-object principle – as argued above.

²¹¹ Although it is not always entirely clear the McFetridge is wrong to interpret Hare in this way. Hare often talks about applying our moral principles to 'all logically possible worlds' (e.g. Hare (1981), p. 113), but seems interested here only in the relatively farfetched examples that sometimes enter moral reflection – e.g. can we torture nuclear terrorists? – and not just *any* possible world.

²¹² Hare (1984), p. 10-11.

poisoning. The laws of nature might have been different. On the other hand, I am not merely claiming that all cyanide that is *actually* ingested causes poisoning. That is, the counterfactual conditional holds true too such that if I *had* ingested cyanide yesterday, I *would* have been poisoned. Thus, causal laws are such that the consequent holds true in nearby possible worlds in which the antecedent (counterfactually) holds true.

By analogy, we seem interested only in a *restricted* range of counterfactual possible worlds, when we defend our moral principles in normative debate. For example, as pointed out by Hare, we ought always to consider possible worlds where we have changed roles with our fellow humans. It is plausible to claim that it is part of using moral concepts that we are willing to engage in hypothetical thinking – not only, but perhaps especially, of the role-reversing kind. That is why Hare's intuition, I argue, is so difficult to dismiss. But there is no need to consider every faraway possible world – possible worlds, for instance, in which the conditions of human life have changed radically or where human sensibilities are radically different. Pace McFetridge, the universal justificatory principles are contingent, though they must be applicable in a restricted class of hypothetical scenarios. McFetridge is, thus, overestimating how strong a principle of supervenience Hare's trustworthy intuition can sustain. All that it will give us is:

Necessarily, if *a* has moral property *M*, then there is a true universal (subjunctive) principle: $(\forall x) (Nx \square \rightarrow Mx)$, and a particular naturalistic truth, *N_a*, such that *Ma* is entailed by these.

And, crucially, this principle, though slightly stronger than (S), is weaker than (XXWA).

Finally, we can also make sense of Hare's intuition in a way that fits the minimalist's broader picture. That is, from the explanatory perspective, it makes sense for two reasons. First of all, allowing for hypothetical tests of moral principles obviously broadens the opportunity for moral debate. If disputants can offer counterfactual counterexamples to each other's moral principles in addition to actual ones, moral debate – and the drive to reach consensus – is significantly strengthened. Also, this is a way for the moral community to plan ahead how to react to possible complex situations in the future. (The ethical problems posed by future bio-

technologies that are already discussed today provide examples of the need for such contingency-planning).

Secondly, we are interested in other people subscribing to subjunctive universal principles because we are interested in their being disposed reliably to judge in accordance with the principles. The problem with Hare's maverick moraliser is that he *appears* to be un-reliable. We cannot be sure, it seems, that he is going to apply the principle the next time it actually becomes relevant, since he appears not to be genuinely disposed to judge in accordance with it. In particular, we are not sure that he will judge in the right way in a future actual situation in which roles are reversed and *we* are on the receiving end of bad consequences.

So, with this explanatory naturalist perspective on moral discourse, we can make a case for the claim that it is implicit knowledge of a principle like (S) – perhaps strengthened to a law – that is partly constitutive of competence with moral concepts. The very function of moral practice suggests that moral consideration is an intra-world business – focussed on the people with whom we are actually involved – though perhaps with a license to visit nearby possible worlds in the course of hypothetical moral thinking.

5. *Weak inter-world moral-on-natural supervenience (McFetridge).*

To finish the discussion of substantive moral-on-natural supervenience, I want finally to investigate whether supervenience-theses stronger than (S) impose moral-on-natural disputational supervenience. Of course, if the argument of the previous section is good, the following two sections are made redundant. That is, if we can establish that (S) (or something law-like and slightly stronger) is true of *actual* moral discourse, there is no need for the minimalist to worry about whether any stronger substantive moral-on-natural supervenience-principles impose moral-on-natural disputational supervenience. However, let me in this and the next section, for the benefit of those who do not yet share Hare's (and my) intuitions about (S), argue that neither McFetridge's nor Jackson's proposed stronger substantive supervenience principles threaten to impose moral-on-natural disputational supervenience.

McFetridge argues that a weak inter-world supervenience principle – namely as mentioned (XXWA) – allows us to conclude that moral truths are *strictly implied*

by naturalistic facts – contradicting Blackburn's (L) above. Could this help to impose moral-on-natural disputational supervenience? McFetridge's preferred supervenience-principle in the moral case is this:

$$(XXWA) \quad (\forall w)(\forall x)[(\forall N) (Nxa \leftrightarrow Nxw) \rightarrow (\forall M) (Mxa \leftrightarrow Mxw)]$$

Thus, according to McFetridge, if some object in the actual world is naturalistically exactly like it is in some possible world, w , then in w it is morally exactly like it is in the actual world. I have already discussed McFetridge's attempt to justify this claim by reference to Hare's intuition above. But let me now focus on what follows from it. He points out that this principle coupled with the claim that any natural object can be fully described by a maximal naturalistic property, N^* , as described above (see NMA) goes to prove that all moral truths are *strictly implied* by naturalistic truths.

To see this, assume for reductio that some moral truth is not strictly implied by any naturalistic truth – in effect, this is the assumption that (L) holds true. This means that, if x has moral property M and naturalistic maximal property N^* in the actual world, then there is a possible world in which x has N^* , but does not have M . But, applying now the supervenience-principle above, (XXWA), this cannot be. The situation in which an object in the actual world shares its maximal naturalistic property with an object in another possible world, yet where they diverge in their moral properties, is exactly the situation that is ruled out by this principle. (Notice that it is exactly the *inter-world* feature of (XXWA) that does the trick). So, we have an absurdity and can deny our assumption. Hence, given (XXWA) and (NMA), we can negate (L) and, thus, conclude that all moral truths *are* strictly implied by natural truths. The conclusion for McFetridge is hence the following necessitation principle, N_{MF} – the negation of (L):

$$N_{MF} \quad (\forall x)(\forall M) [(Mxa \rightarrow (\exists N^*) (N^*xa \ \& \ (\forall w) (N^*xw \rightarrow Mxw)))]$$

Thus, given (XXWA) and (NMA), we know that any moral predicate that applies to an actual object is such that there is a naturalistic description of the actual object from which it is strictly implied that the moral predicate applies to it.

Does this result impose moral-on-natural disputational supervenience? It would seem that not. To see this, imagine again that A believes that Mxa and N^*xa

and B believes that not-Mxa and N*xa. This is the kind of dispute that is not rationally intelligible, if moral discourse disputationally supervenes on naturalistic discourse. The question is then: Could A and B be rationally intelligible while having this moral dispute, given that both A and B are aware of the conceptual truth (XXWA) and the relevant instance of N_{MF}?

It follows from the relevant instance of N_{MF} that A, given that he believes that Mxa, is rationally committed to holding that x has N* and x's having N* strictly implies x's having M. It is, of course, consistent, and rationally intelligible, for A to hold further that x has N*. However, B is not – since he does not believe that Mxa – required to hold that some natural property N* is such that x has N* and x's having N* strictly implies x's having M. Thus, he is not committed (via contraposition) to holding that the falsity of Mxa strictly implies the falsity of N*xa. Hence, he can happily, and rationally intelligibly, endorse N*xa while denying Mxa. Thus, A and B are in no way rationally required to disagree over whether x is N*.

In short, it is not the case that (XXWA) imposes moral-on-natural disputational supervenience.

6. *Strong inter-world moral-on-natural supervenience (Jackson).*

Frank Jackson tries to go one better than McFetridge. He argues that an even *stronger* form of supervenience holds in the moral case and that this allows us to conclude that any moral predicate is necessarily co-extensional with some descriptive (i.e. natural) predicate.²¹³ He even goes on to claim that moral *properties* are natural properties. In his own words, 'the special nature of the famous supervenience of the ethical on the descriptive [natural] shows that any ethical way of carving up the way things are is equivalent to some descriptive way of carving up how things are'.²¹⁴ Does this amount to showing that substantive moral-on-natural supervenience results

²¹³ Jackson prefers to use the term 'descriptive' rather than 'natural' here, though there appears to be no reason why we cannot use the latter – at least in the context of moral discourse. In the following, I shall simply use the word 'natural' in my discussion of Jackson even though he would have used 'descriptive'. For the minimalist, 'descriptive' is a problematic term to use since, given that he is a moral cognitivist, he would accept that our moral vocabulary is descriptive.

²¹⁴ Jackson's argument is in his (1998), p. 122-123.

in disputational moral-on-natural supervenience? Jackson's very strong version of the moral-on-natural supervenience is this:²¹⁵

(S_J): For all possible worlds, w and w^* , if w and w^* are exactly alike descriptively then they are exactly alike ethically.

This is a *global* supervenience thesis in Jackson's terminology in that it quantifies over worlds, rather than over individuals within worlds. He uses this version because he wants 'complete ways things might be' to form the supervenience-basis. This, in turn, is necessary to avoid that variation in things outside an object could play a role in determining its properties of the supervening kind. (He appears to be thinking especially about the psychological case where a semantic externalist would allow that the subjects environmental conditions determine intentional content – an irrelevant consideration in the ethical case). This, I assume, is another way of achieving what McFetridge achieved with his (NMA) – that we are working with maximal naturalistic descriptions of objects.

Jackson can therefore, as he himself acknowledges, say what he wants to say using only 'the appropriate *inter-world* supervenience-thesis'.²¹⁶ If we keep in mind that the principle (NMA) holds true – that there is a maximal naturalistic predicate for every object (which presumably includes information about, e.g., whether a subject is in causal contact with H₂O or XYZ) – then it is clear that the corresponding inter-world, cross-object thesis should serve all Jackson's purposes. Thus, Jackson-supervenience comes down to the strongest possible supervenience claim stating that if two objects inhabiting different possible worlds share all natural properties, then they share all moral properties as well:

$$(XYWW') \quad (\forall w)(\forall w')(\forall x)(\forall y) [(\forall N)(N_{xw} \leftrightarrow N_{yw'}) \rightarrow (\forall M)(M_{xw} \leftrightarrow M_{yw'})]$$

²¹⁵ Jackson (1998), p. 119 and Jackson and Pettit (1996), p. 84. As will become evident below, this principle differs from McFetridge's principle in two ways. Firstly, it is a cross-object principle. And, secondly, it is an inter-world principle with no restrictions tying us to the actual world. We must compare objects in all possible worlds with objects in all other possible worlds, not just objects in the actual world with objects in all other possible worlds. Or more precisely, McFetridge's principle allows – and Jackson's principle rules out – that an object in some *non-actual* world is naturalistically identical to itself in another *non-actual* world, yet that they differ in moral status, as long as the moral predicate in question does not apply to any object in the actual world.

²¹⁶ Jackson (1998), p. 10n.

With this in place, we can now with Jackson define a sentence E as being about ethical nature in the sense that it satisfies the following three clauses: (1) E is formulated in ethical and naturalistic vocabulary. (2) Every world in which E is true has some ethical nature. (3) For all worlds w and w' , if E is true at w and false at w' , then w and w' differ ethically. (Hence, E must not be a disjunction with a moral and a non-moral disjunct that is true due to the truth of the non-moral disjunct). And we can then follow Jackson in noticing that any world at which E is true will be a world that can be described in naturalistic terms too. There cannot be a moral fact that does not also involve a naturalistic fact – an act that is wrong, say, which is not also an act that is an instance of lying or killing or something else naturalistically descriptive.

Now, take E to be the particular claim that a is right – or Ra . Let w_1, w_2, \dots (possibly ad infinitum) be the possible worlds in which Ra is true and let N_1^*, N_2^*, \dots (possibly ad infinitum) be the maximal naturalistic descriptions of a in these possible worlds. Then an equivalence of the following form must be necessarily true:

$$(M) \quad Ra \leftrightarrow N_1^*a \vee N_2^*a \vee N_3^*a \dots$$

Call the disjunction on the right-hand side \underline{N}^* . The proof of this claim is as follows. In every possible world in which it is true that a is right, at least one of the disjuncts on the right-hand-side is true. I.e. Ra strictly implies \underline{N}^* . And, every possible world in which one of the disjuncts in \underline{N}^* is true will be a world in which it is true that a is right. Otherwise there would be two objects (perhaps in different possible worlds) that were maximally naturalistically identical, yet they differed with respect to moral properties and this would contradict (XYWW'). That is, \underline{N}^* strictly implies Ra . The proof of this mutual strict implication ensures that the equivalence above is necessarily true.

(Notice that McFetridge's supervenience-thesis, (XXWA), is too weak to support this argument. In particular, (XXWA) does not support the proof of the right-to-left conditional, since this principle allows that an object in some *non-actual* world is naturalistically identical to itself in another *non-actual* world, yet that they differ in moral status, as long as the moral predicate in question does not apply to any object in the actual world. Of course, intra-world supervenience-theses are also too weak to sustain the argument).

Jackson's conclusion can now be generalised to apply, not just to singular moral propositions involving the predicate 'is right', but universally and to any moral predicate. Given (XYWW') and (NMA) and given that we have above a method for giving the truth-conditions of *any* particular moral proposition (involving any particular moral concept, M, and any particular object, x) in naturalistic terms, Jackson can show that for every moral predicate, M, there is a (highly complex) naturalistic predicate, N, such that these predicates are necessarily co-extensive, or:

$$N_J \quad (\forall M)(\exists N)[(\forall w)(\forall x)(Mxw \leftrightarrow Nxw)]$$

Jackson draws the further conclusion from this that 'moral properties are descriptive [natural] properties'. He argues for this further inference by arguing that we should work with a concept of properties as 'properties-in-nature'.²¹⁷ With this conception, we can rule out that logically equivalent predicates can pick out different properties. But, we are not committed to grant Jackson this conception of a property. We might argue instead that we ought to work with a conception of properties that sees properties as closely linked to our predicates and concepts. And we might hold that the criterion of identity for concepts says that two concepts are identical if and only if one cannot think of the one without thinking of the other. In that case, since we can think using one concept without thinking with another *co-extensional* concept – certainly this is possible in the example of *Ra* and *N** above – we must allow for co-extensive concepts that pick out different properties.

So, let me stay with N_J above – ignoring Jackson's identity-claim – when I now ask whether (XYWW') can impose moral-on-natural disputational supervenience. Clearly, we have to stretch the notion of rational commitment very far in order to assume that rationally intelligible moral deliberators have drawn the very un-obvious consequence of Jackson's – in its own right both un-obvious and controversial – substantive moral-on-natural supervenience thesis that the principle N_J is true. But, let us see what happens if we assume that all rational moral disputants recognise the truth of (XYWW') and its consequence N_J .

Imagine again that A believes that Mxa and N^*xa and B believes that not- Mxa and N^*xa and that A and B are both rational. Now, this assumption of failure of

²¹⁷ Jackson (1998), p. 16 and pp. 125-28.

disputational supervenience leads to a contradiction given that rational people recognise that N_J is true and know the relevant instance of the principle – i.e. $(\forall w)(\forall x) (Mxw \leftrightarrow N^*xw)$. A, given that he believes that Mxa , is rationally committed to believing that N^*xa . And he does believe this and is accordingly in the clear. However, B is irrational, since he is committed (via contraposition on the instance of N_J) to holding that the falsity of Mxa strictly implies the falsity of N^*xa and, hence, he holds that N^*xa is true while being committed to holding that it is false. So, we have a contradiction, since we assumed that B was rational, and we can conclude that failure of disputational supervenience is impossible given N_J .

In other words, the strong inter-world supervenience thesis appears to impose moral-on-natural disputational supervenience. However, let me attempt to resist this conclusion by making a point about the naturalistic predicate N that figures in Jackson's equivalence. What Jackson appears to show is that whenever it is possible to make a distinction between possibilities using a moral concept, it is possible to make the same distinction using only a complex naturalistic concept. However, what he does not show – as he himself readily acknowledges²¹⁸ – is that it is possible for humans to *acquire* that naturalistic concept, that it is *learnable*.

The fact that we know that there will be some possibly infinite disjunction of naturalistic concepts that will capture the extension of any moral concept does not ensure that there will be some single naturalistic concept mastery of which is both possible for human beings and is such that it will enable the subject to 'go on' in exactly the same way as mastery of the moral concept would enable him to 'go on'.²¹⁹ Infinite disjunctions are not learnable. Thus, Jackson's argument from strong supervenience to N_J does not show that moral vocabulary is dispensable in practice and it does not guarantee that moral concepts are co-extensive with naturalistic concepts *that we humans can grasp*.

How is this point going to change the fact that Jackson appears to be able to show that moral discourse disputationally supervenes on naturalistic discourse? Well, we can notice that the above proof of this worked by showing that A and B had – if they wanted to be rationally intelligible – to disagree about the truth of N^*xa where N, we now know, is a hyper-complex naturalistic concept that it is potentially

²¹⁸ Jackson (1998), p. 124-25.

humanly impossible to grasp. But, then, if we are to make the argument for disputational supervenience work, we have to require of A and B that they disagree about some natural fact that they cannot even conceptualise or be irrational. And then we may ask whether our notion of a rational commitment ought to make us deem someone irrational just because he does not disagree about something that he cannot even in principle grasp. The notion of a rational commitment has already been stretched very far by rationally requiring that all un-obvious conceptual truths have to be recognised by any rational deliberator, but this certainly seems to be stretching it too far. It would be to require the impossible of rational moral deliberators!

So, Jackson's strong supervenience principle does not – despite the fact that it imposes N_J – help to establish disputational supervenience. And, to remind ourselves, Jackson is arguably anyway wrong about what actual substantive moral-on-natural supervenience amounts to, as was argued above. So, in conclusion, there appears to be no threat to the disputational purity of moral discourse from even strong forms of substantive moral-on-natural supervenience,

7. Substantive naturalism and disputational supervenience.

I turn now to an investigation of the prospects for some form of substantive – or in McDowell's terminology 'bald' – naturalism to enforce moral-on-natural disputational supervenience and, thereby, moral Cognitive Command. Let me start out by registering a general intuitive scepticism about the 'bald' naturalist's project. The explanatory approach to moral discourse discussed in the previous chapter allows us to see moral discourse as a language-game with a distinct function in our lives – roughly the search for a normative consensus ensuring social co-ordination and peaceful co-operation, including a dynamism in the common system of norms. Crucially, this is a function different from the one served by, say, discourse about our immediate surroundings or the discourse of the natural or social sciences. But, if one adopts this perspective on human discourse – call it a Wittgensteinian perspective – it will from the outset seem misguided to attempt to link up concepts from different discourses in the manner of the substantial naturalist. Why *should* attempts to find

²¹⁹ This point, I take it, is also made by McDowell in his (1981).

such inter-discourse link-ups be successful given that different discourses have evolved exactly to do different jobs? Is the idea that naturalistic discourse could somehow take over all the functions of normative debate – recover all the features of normative debate within a naturalistic language? If that is the idea, it certainly seems doubtful that the ‘bald’ naturalist will be successful and the minimalist’s approach – to respect the autonomy of the different discourses – seems much more wise.

However, the attraction of ‘bald’ naturalism is undoubtedly that it would appear that if moral discourse can be shown to be ‘really just’ naturalistic discourse in disguise, then the objectivity of naturalistic discourse would carry over to moral discourse. Witness Peter Railton’s conception of his own project:

‘The possibility I sought to explore is one in which methodological naturalism might lead – via a synthetic identity claim – to an a posteriori vindication of the cognitive aspirations of moral discourse, without thereby rendering moral judgements systematically false, as they are according to ‘error-theories’.’²²⁰

In other words, it is claimed that the naturalist will save moral discourse from forms of antirealism and non-objectivism – including failure of Cognitive Command. In the remaining part of this chapter, however, I want to argue that ‘bald’ naturalism does not threaten to bestow objectivity on moral discourse in this way. Though one kind of analytical naturalism does appear to threaten to impose moral-on-natural disputational supervenience, it has its own seemingly insuperable problems. And non-analytical forms of naturalism do not impose moral-on-natural disputational supervenience, I argue. Hence, the objectivist hopes for moral Cognitive Command based on the prospects of ‘bald’ naturalism – as described by Railton above – will be undercut by the argument presented in this chapter.

The plan for the discussion of the two forms of substantive naturalism – analytic and ontological – is as follows. In the remainder of this section, I distinguish the different kinds of naturalism. In the next section, I, then, argue that forms of *analytical* naturalism appear to impose disputational supervenience, (at least when we work with strict notions of rational commitment). I look into the suggestion that any form of analytical naturalism is disqualified by some version of the open question argument. I conclude that a sophisticated kind of analytical naturalism – a kind based

²²⁰ Railton (1993b), p. 316.

on network-analysis – does *not* fall prey to this famous argument (section 8). But, I also argue that the version that survives the open question argument will not succeed in securing moral Cognitive Command through imposing disputational supervenience (section 9). I then turn to *ontological* naturalism, arguing that this form of naturalism does not threaten to impose moral-on-natural disputational supervenience. First of all, there can be no rational commitment to knowing the ontological naturalist's *a posteriori* equivalences. Secondly, there is a serious independent problem for the ontological naturalist to do with his inability to account for the univocality of radical moral disagreements (section 10).

But, first, some general taxonomy of substantive naturalisms. Substantive naturalist theories claim that moral properties *just are* natural properties or that moral concepts are conceptually equivalent to natural concepts – they advance identity- or equivalence-claims. What seems relevant in our current discussion of disputational supervenience is that such identity or equivalence claims involve the claim that moral and natural concepts are necessarily co-extensive. Thus, they involve the claim that, for any moral property M, there is some natural property N such that:

$$(R) (\forall x)(\forall w) (Mxw \leftrightarrow Nxw)$$

This characterisation, however, leave many questions open and, consequently, there are many species of substantive naturalism. One question is whether we can come to know these equivalences *a priori* or merely *a posteriori*. A second question is what natural properties make up the reductive class – the N-properties.

Depending on how the first question is answered, two general types of theory are possible. The claim of the *analytical* naturalists is that philosophical analysis can reveal that the content of moral predicates can be identified with the content of certain (often very complex) naturalistic predicates. Thus, the claim is that we can gain *a priori* knowledge of necessary equivalences like (R) above. Perhaps the most refined current attempt to give such an analysis is Frank Jackson's (and Phillip Pettit's) moral functionalism – a version that I shall discuss later.²²¹

Ontological naturalists, on the other hand, believe that we can come to know principles like (R), but only through *a posteriori* investigation – that is, not through

²²¹ Jackson (1998) and Jackson and Pettit (1988), (1995), (1996).

mere conceptual analysis, but through empirical discovery. We can discover that moral terms are necessarily co-extensive with non-moral terms while not being co-intensional with these. Prominent defenders of this approach are Peter Railton and the Cornell Realists like Sturgeon and Boyd.²²²

Depending on how the second question – about the reductive class – is answered a multitude of theories can be advanced. They can be summarised in three groups:

i) *The subjectivist naturalist*: If the assignment of truth-conditions to a moral judgement involves the use (hidden or overt) of indexical terms like "I" or "we" – where the reference is dependent on the context of making the judgement – the account becomes strongly relativist or subjectivist. A crude form of this could be:

‘x is good’ as spoken by A means ‘x is approved by A (when well-informed and consistent)’

ii) *The intersubjectivist naturalist*: There is also the option of referring to *the judgements* of *all* human or rational beings, under some specified ideal circumstances, in one's analysis of moral judgements. This would be a naturalistic version of dispositionalism. David Lewis provides an example of what such an analysis might look like (though he does not endorse it):²²³

x is a value if and only if all human beings would be disposed to value x under conditions of the fullest imaginative acquaintance with x

iii) *The objectivist naturalist*: The objective naturalist avoids referring to the verdicts of subjects under certain specified circumstances when providing naturalistic truth-conditions for moral judgements – against what we saw under (i) and (ii). Rather, the objectivist naturalist gives a natural state of affairs (perhaps, though, a psychological state of affairs) directly as the truth-condition. An example could be the utilitarian theory of Peter Railton:²²⁴

²²² Railton (1986), (1993a), (1993b), Sturgeon (1985) and Boyd (1988).

²²³ See Lewis (1989).

²²⁴ Let me note in passing that Mackie's queerness objection to Platonist versions of cognitivism does not appear to cause problems for any of these naturalistic theories. If moral states of affairs are identical to natural states of affairs, then they are obviously not queer in the relevant sense. With respect to the epistemological queerness-challenge, the mere identity of moral properties with natural properties does not – unless this identity can be known a priori – ensure that we have a straightforward account of moral epistemology. But, there is at least the presumption that an account that

x is morally right if and only if x maximises the satisfaction of everyone's non-moral interests, considered equally and in a state of full information.²²⁵

But, let me now turn to the question of the remainder of this chapter: Do any of these forms of naturalism impose moral-on-natural disputational supervenience?

8. Analytical naturalism, disputational supervenience and open questions.

The analytical naturalist holds that it is possible to conceptually analyse any moral predicate using only natural predicates. Thus, where M is a moral predicate and N a naturalistic predicate, he claims that:

It is a priori that: $(\forall M)(\exists N)[(\forall w)(\forall x)(Mxw \leftrightarrow Nxw)]$

This means that, according to the analytical naturalist, there are a priori, necessary equivalences of the form: $(\forall x) (Mx \leftrightarrow Nx)$ for every moral predicate. Thus, we must grant that all rational deliberators know such principles, given the strict notion of rationality we have reluctantly accepted. To see, now, whether this imposes moral-on-natural disputational supervenience, we must ask whether a priori knowledge of such necessary equivalences commits rational deliberators to a naturalistic disagreement whenever they have a moral disagreement.

However, we already saw in section 6 – while discussing Jackson's supervenience-principle – that knowledge of instances of the above principle does commit disputants to naturalistic disagreements whenever there are moral disagreements. That is, as long as it is possible for humans to grasp the naturalistic concept that is used to analyse the moral concept – as was *not* the case with Jackson's supervenience-based equivalence – there is no reason to think that such graspable analytical principles do not secure a priori that strictly rational disputants always disagree naturalistically whenever they disagree morally. In other words, if the analysing naturalistic concepts are graspable, the truth of analytical naturalism does threaten to impose moral disputational supervenience. Add to this the fact that there

does not involve reference to extra-sensory faculties can be given – given that the property, we have to identify, is natural.

²²⁵ Railton (1986), pp. 141-155.

are candidates for such graspable analyses on the market and we have the appearance of a problem for moral minimalism. Jackson himself, for instance, goes on to use a network analysis (performed on what he calls mature folk morality) of moral concepts to identify the relevant graspable, though very complex, naturalistic concepts. I shall shortly look into this in more detail.

The minimalist – wanting to block this road to moral objectivity – may at this point return to the starting-point of the dialectic of this chapter and revoke the definition of rational commitment reluctantly given there. It may be argued that demanding of any rationally intelligible deliberator that he knows all – even highly unobvious – conceptual truths was, after all, being too demanding. But, there is still the question whether there are *other* ways of stopping analytical naturalism from imposing moral-on-natural disputational supervenience. I shall discuss two possible ways of doing this. First of all, one might show that any analytical naturalism must fall prey to a refined version of Moore's open question argument.²²⁶ My investigation in the remainder of this section will show, however, that not all forms of analytical naturalism do so fall prey. Secondly, in the next section, I argue that the forms of analytical naturalism that survive the open question challenge have other problems and, thus, cannot help to establish that moral discourse exhibits Cognitive Command. But, first the open question argument.

The open question argument is based on the observation that, for any attempt to define a moral term M in purely non-moral terms, N, it will make sense to ask questions like: But is being describable in these naturalistic terms *really* being M? For instance: Is being maximally conducive to the overall happiness of everyone involved *really* being right? Another way to put this is by noting that it never – no matter what naturalistic terms have been used in the definition – appears to be *self-contradictory* to suppose that something has the natural property, N, but fails to have the moral property, M. But, the argument then goes, if N really succeeded in defining M, then such questions could not intelligibly have been posed and denials of the definition would have been self-contradictory. Hence, no naturalistic definitions of moral terms

²²⁶ Moore (1903), chapter 1, section B.

can be defended. Moral concepts defy all attempts to be analysed using only naturalistic concepts – the project of analytical naturalism is doomed.²²⁷

However, there is an assumption underlying this formulation of the open question argument. The assumption, namely, that the outcome of all conceptual analyses must be obvious and un-informative. Without this assumption, clearly it *would* make sense – in one sense of 'making sense' – to ask questions and denials of definitions *would not* appear self-contradictory. But, as discussions of the Paradox of Analysis show, this assumption is far from true. To see this, let me give an account inspired in part by Michael Smith of what conceptual analysis is – showing why analyses can be both informative and un-obvious.²²⁸

Acquiring a concept is coming to treat certain truths as platitudinous – if not consciously, then at least implicitly in one's practice with the concept. Thus, the platitudes involving a concept are the propositional expressions of a set of inferential and judgemental dispositions – knowledge-that about knowledge-how – the possession of which are necessary and sufficient conditions for possession of the concept. But then, arguably, one has succeeded in analysing a concept, M, in terms of some other concept, N, only if mastering the concept N involves coming to treat as platitudinous all the platitudes involving M. Now, one way to arrive at such an analysis is by summarising and systematising all the M-platitudes in some simple formula. As an example of such a 'summary style' definition, one could mention the dispositional account of the colour red: x is red if and only if standard observers under standard conditions judge x to be red.²²⁹ The claim would here be that the right hand side of this equivalence captures all the relevant platitudes governing our use of 'red'.

But, if this is the correct account of conceptual analysis, then it is also clear why analyses do not have to be un-informative and obvious. It may be that a subject who is competent with a concept – who has the know-how – has nevertheless never acquired the corresponding knowledge-that about the platitudes involving the concept. Hence, an analysis may be *informative for him*. And, it may of course require

²²⁷ Let me notice in passing that the ontological naturalist has a direct answer to the challenge from the open question argument. He is not even engaged in the project of giving a semantic analysis. He merely identifies some moral property with some natural property on *a posteriori* grounds. And, since it is no wonder that it makes sense to ask questions about matters that depend on what is actually the case in our world, the open question argument does not get a grip here.

²²⁸ Smith (1994), p. 35-41.

²²⁹ 'Summary-style-definition' is Smith's term – Smith (1994), p. 38.

a lot of conceptual reflection to see – and, therefore, be highly *un-obvious* – that some particular summary of the platitudes is adequate. (For a simple example of analytical truths that are not immediately obvious – and, hence, make questioning intelligible – consider the products of complex arithmetical calculations).

To make these points about analysis clearer, let me elaborate a little. When an analysis results in the discovery of a necessary equivalence like $(\forall x)(\forall w) (Mxw \leftrightarrow Nxw)$, what has been discovered is that two judgements cannot but share their truth-conditions. That is, what we do when we perform an analysis – like the platitude-analysis suggested by Smith above – is to break down the content (understood as the truth-conditions) of the two judgements into their component parts. What we may then discover, and what the analytical naturalist believes to have discovered in the case of certain moral and natural judgements, is that two judgements actually have the same component parts – i.e. capture the same set of platitudes. This is what ensures that they are made true by the same state of affairs – i.e. that they have the same truth-conditions and that the component concepts refer to the same properties.

But notice also that this is, arguably, not sufficient to ensure that the judgements have the *same sense* (or that they are the *same thought* or that their linguistic expressions *mean the same*). If we accept a Frege-inspired account of meaning and sense, then the sense of a sentence is constituted by *the way we determine* its truth-value. And crucially, since the two judgements conceptualise the same content (i.e. truth-conditions) in *different ways* – or, to use a metaphor, *carve up* the same content in different ways – the ways in which the truth-values are determined also differ. Therefore, given the Frege-inspired account of sense, they constitute different thoughts, even though they have the same truth-conditions. Hence, what the analytical naturalist's analysis (allegedly) establishes is that certain naturalistic judgements must share their truth-conditions with certain moral judgements, not that they have identical senses. They establish that naturalistic and moral concepts refer to the same properties, not that they are identical.

These distinctions allow us to see again why an analysis need not be un-informative and obvious. That is, someone may possess two concepts – a moral and a naturalistic – with different senses. Because of their differing senses, it is quite intelligible that he has not discovered that these different concepts deal with the same

content. He may not yet have broken down the contents of judgements involving these concepts into their constituent parts – i.e. he may not yet have gained reflective knowledge of the platitudes surrounding the two concepts. Therefore, it may not be obvious to and un-informative for him that there is co-incidence in the surrounding platitudes and, hence, that judgements applying the concepts to the same objects have the same content (i.e. truth-conditions). The analysis may quite intelligibly enlighten him about the identity of content. So, in conclusion, a premise in the open question argument – that all analyses are un-informative and obvious – is false and the analytical naturalist would appear to be out of the woods.

Notice now, however, that from the account of conceptual analysis just given it is clear that the open question argument can survive in a different form.²³⁰ That is, we have just learnt that any adequate conceptual analysis must manage to circumscribe *all* the platitudes surrounding the analysed concept. But, then a reinterpretation of the open question argument would be to say that what it points to is the fact that any naturalistic analysis of a moral term must miss a crucial aspect of the use of this term – must fail to capture at least one moral platitude. And, the feature of our use of moral concepts that is so elusive for the analytical naturalist is generally regarded – by those who place emphasis on the open question argument – as being the *normativity* of moral concepts. To accept a moral belief about some situation is always to be guided or directed in some way in relation to this situation. Moral beliefs tell us what to do, give us reasons for acting, recommend a certain course of action, call for a certain 'response of engagement'.²³¹ The reason why no naturalistic analysis can capture this feature of moral practice, it is then claimed, is that merely believing something about the natural features of a situation never by itself can involve such guidance or directedness. This second open question argument is just what has also been named the 'So what'-argument.²³² The point of this argument is exactly that any purely natural description of a state of affairs – say, that this act is the act that ideal deliberators would as a matter of fact desire to perform – always can be

²³⁰ The distinction between the two versions of the open question argument is also found in Jackson (1998), p. 150-53. He calls them 'the Moorean version' and 'the Humean version' respectively.

²³¹ Wiggins' phrase. Wiggins (1993a), p. 331.

²³² See e.g. Johnston (1989), p. 157. And also Darwall, Gibbard and Railton (1997), p. 3-5 and 29. As they remark, Wittgenstein's 'Lecture on Ethics' is an important source for this version of the open question argument.

met with the retort: So what? And this reveals that the proposed analysis is deficient – fails to capture the normativity of moral facts.

How does the analytic naturalist fare faced with this version of the open question argument? Certainly some versions of analytical naturalism are disqualified by it. Think of the crude naturalistic theory that 'x is right' means 'x is what contributes most to the overall happiness of all involved'. The problem with this theory is that from the fact that somebody believes that x contributes to overall happiness, it does not follow that this person takes it that there is a norm commanding x – that he believes that he has a reason to choose x. That is, the crude naturalist fails to capture the normativity of moral judgement in his analysis.

The analytical naturalist can, however, seemingly get out of this trouble caused by the open question argument by including something about desires, pro-attitudes, or rational desires (where 'rational' is naturalistically defined), in his analysis of moral judgement. Suppose that the naturalist suggests the following analysis: 'x is right' means 'x is the property (naturalistically described) that I approve of'. Then, since my belief that x has some natural property that I approve of does give me a (defeasible) reason to choose x, the naturalist has succeeded in capturing the normativity of morals in his analysis. But, now another problem occurs. It is a platitude that if I say that x is right and you deny it, we disagree. In other words, it is a platitude that moral claims aspire to objectivity. But, then our subjectivist analytical naturalist has failed to capture a platitude in his analysis, since I do not necessarily disagree with you that x has the property that you approve of when I hold that x does not have the property that *I* approve of.

So, the second version of the open question argument succeeds in disqualifying all forms of analytical naturalism if any naturalistic analysis of a moral term must fail to capture at least one aspect of our use of the moral term. But, this latter claim would, of course, be a very bold postulate. Why should we exclude the possibility in advance that an all-inclusive analysis of moral terms in naturalistic terms can be given? Indeed, theories have been developed that appear to pre-empt any failure to capture platitudes surrounding the concepts to be analysed. That is, the analytical naturalist appears to be able to avert any problems with open questions once and for all by adopting a sophisticated holistic analytical naturalism – in the form

of a *network analysis* of moral terms. I shall now look briefly at Jackson and Pettit's moral functionalism to see how this is done.

Moral functionalism makes sure that no platitudes are left out in the analysis simply by constructing the analysis directly from the collection of all the relevant platitudes. So, a first question is: How do we identify the totality of moral platitudes? Current folk moral theory is still under negotiation – not yet perfected. On the background of a broad agreement over basic moral principles and platitudes, the functionalist argues, the folk and the philosophers still disagree about the details. Thus, for example, the philosophers (reason-internalists and reason-externalists) disagree about whether it is a conceptual truth that making a moral judgement involves accepting that one has the corresponding reasons. And there are disagreements about how often exactly exceptions from general moral rules are justified and about why they are justified. The various moral theories – Kantianism, utilitarianism, virtue theory and so on – try to give principled answers to questions like these, but no consensus is so far in sight. Thus, we have not yet reached what might be called 'a mature folk moral theory'. But, the moral functionalist argues, it is reasonable to expect that in the future the folk theory will mature.

Mature folk morality, according to the moral functionalist, is a theory that has become internally coherent and consistent as a result of applying the procedures for improving our set of folk moral beliefs that is already there in our current folk morality²³³ – a process that might have an end by achieving something like a final Rawlsian equilibrium. It is this mature theory that identifies for us the moral platitudes that the moral functionalist has to respect in his analysis.

According to Jackson, the platitudes may be divided into three different types.²³⁴ There are what he calls 'input-clauses'. These are principles telling us which non-normative descriptions of a thing warrant an ethical description (i.e. what I have earlier referred to as moral (linking)-principles): 'Lying is wrong except in circumstances c_1, c_2, \dots ' is a simple example. Secondly, there are 'internal role clauses' that link moral propositions with other moral propositions – e.g. principles that show how terms like 'right', 'good', 'courageous' and 'virtuous' can be inter-defined. Included

²³³ Jackson (1998), p. 133.

²³⁴ Jackson (1998), p. 130-31. An example of a list of platitudes that surrounds a particular moral concept – that of being fair – is to be found in Jackson and Pettit, (1995), p. 22-23.

under this heading will also be principles that show how various moral properties weigh against each other and meta-principles like the supervenience-principle. Finally, there are 'output-clauses' – principles that show how moral judgements have implications for the reasons, motivation and, hence, the behaviour of agents. Also under this heading will be general statements about how particular moral considerations give stronger moral reasons and tend to motivate more than other particular moral considerations. This mature theory Jackson calls **M**.²³⁵

The moral functionalist now argues that a Ramsey-Carnap-Lewis-style definition of the moral terms involved in **M** will provide us with a naturalistic analysis. Crucially, we will know that such an analysis is adequate – takes into account all the platitudes – since it is a mere construct from the totality of platitudes. Let me briefly summarise how this is supposed to work. The first step in creating a Ramsey-definition for a moral theory, **M**, is to reformulate it such that all moral predicates are written as a name of a property. So, 'is right' becomes 'has the property of being right'. The second step is to replace all moral terms – now property-names – in **M** with different variables such that we will get an open expression of the form **M**(x_1, x_2, \dots). The third step is then to define the Ramsey-sentence of **M** as:

$$(\exists x_1) (\exists x_2) \dots \mathbf{M}(x_1, x_2, \dots)$$

The fourth step adds a uniqueness clause such that we get the Modified Ramsey-sentence of **M**:

$$(\exists x_1) (\exists x_2) \dots [\mathbf{M}(x_1, x_2, \dots) \ \& \ (\forall y_1) (\forall y_2) \dots (\mathbf{M}(y_1, y_2, \dots) \leftrightarrow x_1 = y_1 \ \& \ x_2 = y_2, \dots)]$$

The final step then implicitly defines the individual moral properties like rightness in the following way:

$$(R_j) \quad A \text{ is right} \leftrightarrow (\exists x_1) (\exists x_2) \dots [\mathbf{M}(x_1, x_2, \dots) \ \& \ A \text{ has } x_r \ \& \ (\forall y_1) (\forall y_2) \dots (\mathbf{M}(y_1, y_2, \dots) \leftrightarrow x_1 = y_1 \ \& \ x_2 = y_2, \dots)]$$

²³⁵ Thus, in contrast to the minimalist, the functionalist treats substantial moral linking-principles as platitudinous – as a priori knowable and necessary. He writes substantial moral principles into the meaning of moral terms,

To paraphrase: we can identify being right with having the (unique) property that plays the rightness-role as it is described on the right-hand-side of (R_J).^{236,237} The right-hand-side of (R_J) does not contain any moral claim whatsoever, since we stripped **M** of moral property-names at an earlier stage in the procedure. Of course, this in itself does not ensure that the analysis given by the right-hand-side is naturalistically reductive, since the range of the quantifiers on the right-hand-side may contain non-natural objects. However, according to Jackson and Pettit,²³⁸ the independent argument from strong substantial moral-on-natural supervenience to the necessary identity of moral properties with natural properties – that I discussed in section six – guarantees that there are no such non-natural objects in the domain. Thus, we are ensured that the properties that are interconnected in the way specified by the right hand side of (R_J) are themselves natural properties and, therefore, that the analysis is reductive.

However, I have already argued that the argument from strong supervenience to the necessary co-extensionality of moral and natural predicates does not go through. As we saw, it presupposes, controversially, that we can make sense of the idea of a concept that is possibly an infinite disjunction. Thus, Jackson and Pettit cannot avail themselves of the conclusion of this argument in their argument here. And then it is not clear that the range of the quantifier does not include non-natural objects – i.e. that the analysis given in R_J is reductive after all. But, let me for now

²³⁶ (R) is knowable a priori. Knowing **M** is a conceptual matter and all the steps above are purely conceptual. It should, however, also be noted that if we want to identify the moral property of rightness with a *specific* non-moral property, we will have to find out what non-moral property plays the role described in the rhs of (R) *in our world* – the realiser-property. But, this cannot be an a priori question. So, to pick Jackson's example, specific identity-statements between moral and non-moral properties like: 'Rightness = maximising expected hedonic value', can only be known a posteriori.

²³⁷ According to Jackson, the modal status of (R) depends on whether it is seen as an instance of meaning-giving or reference-fixing. On the one hand, if the right-hand-side of (R) *gives the meaning* of the left-hand-side, then whatever property plays the rightness-role in a particular possible world will be the rightness-property in that world. 'Right' will have the meaning of a flexible definite description – that of being the property that plays the rightness-role. Thus, every possible world in which lhs is true, rhs is true and vice versa. That is, (R) becomes necessarily true. On the other hand, if rhs *fixes the referent* of lhs, then the equivalence will only be true in a possible world if the property referred to on the rhs in that possible world, is the property that rhs refers to in the *actual* world – and that is a contingent matter. Hence, (R) will come out as contingent. On this construal of (R), the rightness-property becomes what plays the rightness-role in the actual world. Jackson remains agnostic about whether folk morality treats moral terms as rigid or flexible designators. He even suggests that there may be no determinate answer to questions about rigidity of the terms in natural language. Therefore, his moral functionalism remains agnostic about the modal status of reductive equivalences like (R) – Jackson (1998), p. 144.

²³⁸ See Jackson and Pettit (1996), p. 84-85, for this argument.

ignore this problem for the moral functionalist and give the moral functionalist some more rope to see how he unfolds his argument.

If we grant the moral functionalist the validity of the argument we have just dismissed, then he appears to have provided us with a reductive naturalistic analysis of the moral term 'right' – and it would appear that it is immunised against the objection that it does not capture all relevant platitudes involving the term. In particular, it would appear that the moral functionalist can account for the normativity of moral judgement. That is, one of the platitudes making up **M** was the conceptual truth that if it is right for S to ϕ in circumstances, C, then S has a reason to ϕ in C (or it would be rational for S to desire to ϕ in C).²³⁹ If A is right, then A has the property that plays the rightness-role as specified on the right hand side of (R_j). But then, part of playing the rightness-role is to be the property that it is rational for one to desire. In other words, whenever I believe that A has the naturalistic property that plays the rightness-role, I also believe that A is what it is rational for me to want. And, to believe that something is a rational goal is, of course, to be guided or directed. Moral functionalism, in other words, succeeds in capturing the normative element of moral judgement. And, similar arguments could be produced to show that the functionalist analysis captures all the other platitudes.

Now, we may already have found problems with this general strategy for providing a reductive naturalistic analysis of moral terms.²⁴⁰ But, given that it has the advantage over other versions of analytical functionalism that it is immunised against failure to capture all platitudes, it is perhaps fair to say that it is *the analytical naturalist's best shot at a successful reductive analysis*. So, when I now go on to deal a final blow to the suggestion that analytical naturalism imposes moral-on-natural disputational supervenience, I am sure to be dealing with the strongest version of analytical naturalism when I take moral functionalism as a representative.

²³⁹ That this *is* a conceptual truth is of course contested by reason-externalists.

²⁴⁰ Smith argues that there is a permutation-problem. An assumption is made when the uniqueness clause is added that there will be exactly one realisation of the network of commonplaces that makes up **M**. That there is exactly one (natural) property that plays the rightness-role, to take the example from above. But, what ensures that this assumption is true? Is it certain that the set of platitudes are rich enough in content to pick out exactly one natural property for each moral term? Smith (1994), p. 48-52.

9. *The prospects for analytical naturalism.*

With these preliminaries in place, I now want to reach a conclusion about whether moral functionalism poses a serious threat to moral minimalism. The question is therefore immediately: Can the moral functionalist – possibly the only analytical naturalist that survives the open question argument – help to establish that moral discourse disputationally supervenes on naturalistic discourse and, therefore, stands to inherit its objective status? The case for a positive answer is straightforward. The moral functionalist's claim is that an equivalence can be established *a priori* between any moral concept and some non-moral concept in the shape of principles like (R_J).²⁴¹ Thus, to take Jackson's example of the concept of rightness, every strictly rational moral deliberator who has acquired the concept of moral rightness will know a priori that (where M' is moral rightness and N' is the complex rightness-role-property figuring on the right-hand-side of R_J):

$$(R') \quad (\forall w)(\forall x) (M'xw \leftrightarrow N'xw)$$

But, then it might appear to follow that it is a priori that any disagreement about whether something is right must involve a disagreement about whether something has the natural property that plays the rightness-role. Is this, then, proof that moral discourse disputationally supervenes on naturalistic discourse?

It would seem so. Imagine that A believes that x has M' and B believes that x does not have M' – i.e. A and B disagree morally. Given the demand of strict rationality, A and B both recognise (R'). But then A ought to believe that x has N' and B ought to believe that x does not have N'. Thus, A and B have a rational commitment to a further naturalistic disagreement. Hence, transmission of objectivity from naturalistic discourse to moral discourse appears to go ahead.

However, we ought at this point to recall that there are many obstacles to overcome for the moral functionalist in order to reach this conclusion. That is, at least two places in the previous argument are places where it seems possible to block the case for moral objectivity based on the defence of a moral functionalism. (1) The moral functionalist must heavily exploit the early concession we gave to the analytical

naturalist. Namely the concession that we should work with a strict notion of rationality. Since the case for making this concession was very weak, we should remember that so is the analytical naturalist's case for employing it in his argument.

But, (2), it has also been shown that Jackson's (and Pettit's) use of the argument from strong supervenience to the necessary co-extensionality of any moral predicate with some natural predicate is flawed in two ways. First of all, it was based on the un-argued and false assumption that actual substantive moral-on-natural supervenience is of a strong inter-world kind. Secondly, the argument for co-extensionality itself is also shaky, since the analysing naturalistic predicate was shown to consist in a potentially infinite – and hence un-graspable – disjunction. So, there appears to be no way of defending the moral functionalist's appeal to the argument from substantial supervenience to identity. But as already mentioned Jackson and Pettit need this argument to ensure that their analysis in the shape of (R_J) is really reductive. Without it there is no guarantee that the right-hand-side quantifiers do not range over non-natural objects. In other words, the argument for the pivotal claim of moral functionalism – namely that (R_J) constitutes a *reductive* analysis – is very shaky indeed.

But, as if this was not already enough to show that the odds are against the moral functionalist, I now want to make an additional point against him. Ultimately, this argument places the moral functionalist in the same boat as the strong moral relativist discussed earlier. According to the moral functionalist, we can come to know principles like (R') through performing a certain procedure on mature moral theory, **M**. But, according to Jackson, we do not have any guarantee that there will be convergence on a single set of moral platitudes – no guarantee of convergence on what principles belong to **M**: 'Indeed, I take it that it is part of current folk morality that convergence will or would occur.....But, this may turn out to be, as a matter of fact, false'.²⁴² This, we might misinterpret as an admission that there is no a priori guarantee that all moral disagreements will involve a cognitive shortcoming – i.e. as a statement of failure of Cognitive Command. However, this is not what Jackson's moral functionalist would have to say. In fact, he appears to be able to argue that

²⁴¹ If moral terms are not read rigidly, the moral functionalist argues, these are necessary equivalences too.

²⁴² Jackson (1998), p. 137.

there *is* an a priori guarantee that moral disputes always involve cognitive shortcomings.

Since, namely, the moral functionalist holds that *substantive moral principles are written into the meaning of moral terms* – we saw this when we listed the platitudes that according to Jackson make up mature moral theory – he can rule out a priori the very idea of a radical moral disagreement. This point I made earlier in the discussion of Tersman's objection in chapter three. If substantial principles are included in the meaning of moral terms, any evidence of a radical moral disagreement is evidence that the disputants are using moral words with different senses. Or, in Jackson's words: '...if two people agree on the descriptively given facts, are not confused, and one uses 'right' to describe a given situation and the other does not, they mean something different by the terms – or at least they do on one reasonable meaning of 'meaning'.'²⁴³ Thus, if a radical moral disagreement appears to persist even between two subjects that are not confused and are correctly informed about the non-moral facts, then they are simply not speaking the same language. There can only have been the *appearance* of a radical disagreement in the first place. Or, to use Price's term, all radical disagreements evaporate. But, then the moral functionalist must argue that the subjects really belong to different moral parishes – in other words, he must ultimately defend a parochialist theory of morals and face all its problems.

So, to re-iterate now points made in earlier chapters, this a priori guarantee comes only at a considerable cost. It results in the failure to accommodate the univocality-intuition – the intuition discussed in chapter three that radical moral disagreements are univocal. But, then, as we saw in chapter five in the course of arguing against strong moral relativism (or parochialism), adherents of this view seems to forget what is an important function of moral discourse – a function that we identified from the explanatory naturalist's perspective. Moral truth-governed debate serves the purpose of allowing people with initially different basic moral outlooks to engage in a process of moral argument that may ultimately lead to the creation of a moral consensus. Saving moral discourse from failure of Cognitive Command by ruling out a priori univocal radical moral disputes is to make this function of moral debate impossible.

²⁴³ Jackson (1998), p. 161.

Now, as a final attempt to defend the moral functionalist here, it might be protested that constructive debate across parishes will still be possible – even on the moral functionalist’s account. That is, subjects speaking different moral languages might still come together *to discuss what language to speak* – what moral concepts to use. They might so to speak semantically ascend to a higher level of moral debate and then perhaps still achieve the purposes of moral debate – that is, achieve cross-parish social co-ordination and dynamism in the sets of moral norms. However, I do not see how this is going to help. Surely, if two subjects from different parishes were to discuss the issue of what moral concepts to adopt, they would have to start by discussing what moral concepts they *ought* to adopt – where this ought would be a moral ought. But, then they would be back to square one. And the issue about univocality would re-emerge. The point here is simply that there is no way of escaping in this situation the normative perspective that one already adopts.²⁴⁴ That is why an account – like the minimalist’s – that allows for univocal radical disagreement (in the already specified sense) is to be preferred to the analytical naturalist’s account.

Finally, let me repeat that there is no problem with the fact that the above argument against analytical naturalism depends crucially on features of one particular kind of analytical naturalism – namely moral functionalism with its use of **M** as a basis for analysis. We here need to keep in mind what was argued in the previous section. Moral functionalism is probably the strongest possible variant of analytical naturalism in that it appears to be immune to the problems posed by refined versions of the open question argument. In fact, it is exactly the feature of moral functionalism that is used to run the above argument against moral functionalism – the use of **M** as a basis for analysis – that saves this theory from the open question argument. So, it would appear that the argument to the conclusion that there is no threat from analytical naturalism to minimalism covers all *viable* forms of analytical naturalism.

²⁴⁴ Remember here again a point from the discussion of parochialism in chapter five. If the two subjects try to discuss the matter of what moral concepts to adopt without making it into a normative discussion, then whatever the outcome of the debate turns out to be, it would not have the normativity required for it to have categorical force for the subjects.

10. *Ontological naturalism and disputational supervenience.*

The question remains whether *ontological* naturalism can impose moral-on-natural disputational supervenience. Can the a posteriori discovery that moral predicates are necessarily co-extensive with some natural predicate help establish that moral discourse disputationally supervenes on naturalistic discourse? We might interpret Railton in the passage quoted in the beginning of this chapter as claiming something like that. However, I shall argue that there are reasons to answer this question in the negative and, further, to hold that some versions of ontological naturalism lead to an indefensible relativism. But, first, a brief introduction to ontological naturalism.

The ontological naturalist starts out by claiming that there is a similarity between the way moral terms function and the way proper names and natural kind terms function according to Kripke (and others). That is, they can all be seen as functioning like rigid designators. According to the Kripkean view, the term 'water', for example, is a natural kind term that rigidly designates whatever it is that is causally responsible for our use of the word in the actual world – in the same way as we may believe that a name designates the same person in every possible world that it designates in the actual world. Now, 'water' refers to the actual watery stuff of our acquaintance – which we have, of course, discovered to be H₂O. But, water, then, will be H₂O in every possible world (in which there is any water). In which case the claim that the property of being water is identical to the property of being H₂O – and the corresponding claim about co-extensiveness: $(\forall x)(x \text{ is water} \leftrightarrow x \text{ is H}_2\text{O})$ – becomes a necessary truth.

Crucially, this metaphysically necessary relation cannot be discovered through conceptual analysis alone, since it is only knowable a posteriori that the watery stuff of our acquaintance is H₂O. So, when we accept that certain terms refer rigidly, we can expect for there to be *a posteriori* knowable necessary identity-claims involving these terms. Also, it follows that *analysis* becomes impossible. As we have seen, conceptual analysis can be seen as an attempt to capture the a priori, necessary truths whose (perhaps tacit) acceptance is what constitutes the mastery of a concept. But, just as there are no platitudes surrounding the use of a particular proper name, there are no platitudes governing natural kind terms rigidly construed. Truths about the

watery symptoms – that (pure) water is fluid, transparent, tasteless, potable etc. – that we use in this world to fix the reference of 'water' are not true in all other possible worlds in which H₂O figure. Hence, they are not candidates for being a priori, necessary truths about water. The only essential truth that is connected with 'water' is that it is the watery stuff of our acquaintance.

Now, the ontological naturalist attempts to adopt this strategy in the moral case. He rejects reductive *analysis* of moral concepts. The only claim he advances on a priori grounds is that moral terms are rigid designators. He then attempts to find out a posteriori what the natural properties that underlie our use of moral predicates actually are. He believes that he has thereby found a way to be a naturalist – in the sense that he can identify moral properties with natural properties – without falling prey to the open question argument, simply because he offers no semantic *analysis* of moral terms. What he needs to do is simply to study the language community's use of moral language and discover that certain, probably complex, natural properties underlie our (typical) use of certain terms.²⁴⁵ For example, he may discover in this way that the property underlying the use of 'right' in the community is the property of being most conducive to the happiness of all involved.

But, our question is whether some such discovery could ever bestow *objectivity* on morals via the establishment of moral-on-natural disputational supervenience. The answer seems to be a clear no. We have to ask whether – given the truth of ontological naturalism – it is a priori that strictly rational moral deliberators must disagree naturalistically whenever they disagree morally. But, this time we find that the disputants are not required – in order to be strictly rational – to believe in any particular equivalence-claims, i.e. claims like $(\forall x) (Nx \leftrightarrow Mx)$. Equivalence-claims like these are – on the ontological naturalist's account – not knowable a priori, but can be known only through empirical discovery. The only thing that strictly rational deliberators will know a priori on this account is that moral terms function rigidly and, therefore, that it will be possible to know *some* such equivalence-claims a posteriori. The argument against the suggestion that ontological naturalism imposes moral-on-natural disputational supervenience then goes as follows: Only if strictly rational deliberators are committed to believing equivalence-claims like the

²⁴⁵ Railton (1993b), p. 316.

above is there a case to be made for its being a priori that all moral disagreements involve natural disagreements. But, it would be *stretching the concept of rationality* – even strict rationality – *too far to include a commitment to knowing what are mere a posteriori, and furthermore highly un-obvious, truths*. Hence, the argument from ontological naturalism to moral-on-natural disputational supervenience – and further to moral objectivity – is a non-starter. Ontological naturalism is, therefore, unlikely to threaten moral minimalism.

However, I want to add to this – already conclusive, I believe – argument against the ontological naturalist, a further criticism of the very idea that ontological naturalism could help to establish moral objectivity. The problem for the ontological naturalist is that he ends up committed to defending a malignant form of relativism and is troubled by the ensuing loss of univocality – i.e. he is haunted by the problems we saw also affected the analytical naturalist and the parochialist.²⁴⁶ Let me make the point one final time. It is an undisputed fact that current folk moral practice is full of disagreements. To take an example, there is a group in our culture that favours some form of deontological, say Kantian, normative theory whereas another group place their money on a teleological, say utilitarian, normative theory. But, if the ontological theorist investigates these sub-cultures, he is bound to come up with two *different* accounts of what natural property is identical to, for instance, the moral property of being right. Thus, for the Kantians, facts about the intentions of the individual agent will tend to play an important role in determining occasions of correct use of the term 'right', whereas for the utilitarian facts about desire-satisfaction of all agents, say, will be relatively more important.

But, if 'right' is used by the two sub-communities to refer to different properties, there will be no disagreement if one group says that *a* is right and the other denies this. The subcultures *will lose their common subject-matter*. (Another way of seeing this is to notice that moral terms are alleged by the ontological naturalist to function like proper names. And, if, say, the name 'Peter' is used by two different groups to refer to two different persons, then obviously they do not disagree when they ascribe different properties to the two different Peters). Thus, once again, univocality is lost. And this, to reiterate, is to forget that an important function of

moral debate is to make possible the creation of a consensus among disputants that are initially far apart.

The ontological naturalist could perhaps here argue that he is only going to apply his method of finding a naturalistic referent of moral concepts *to concepts as they figure in mature morality*. The ontological naturalist is, then, claiming that the property referred to by the term 'right' in *mature* moral theory is identical to some natural property.²⁴⁷ But this is no solution to his problem. According to the ontological naturalist, we do not know that the moral practices that we engage in today have the potential to mature into one and only one theory – that we have the potential to all converge on morals in the end. Thus, according to him, it is possible that the seeming disagreement in today's folk morals is really a symptom of there not being one single theory in play. Perhaps future refinements of our moral folk theory will actually result in the creation of different moral communities with each their mature theory. So, the ontological naturalist who wants to save his argument by appealing to mature moral theory has no guarantee that there ever will be one unique mature theory. In that case, he has no guarantee against the familiar problem of strong relativism and the ensuing loss of univocality that I have discussed several times in the foregoing.

To sum up, it has been argued that neither of the three candidates for imposing moral-on-natural disputational supervenience – actual substantive moral-on-natural supervenience, analytical and ontological naturalism – has proved successful. Actual substantive moral-on-natural supervenience does not impose disputational

²⁴⁶ I here follow, I think, Smith (1994), p. 32-35. He, in turn, is inspired by Hare. Recently, Horgan and Timmons (1992) have advanced similar arguments.

²⁴⁷ I believe Smith (1994) is someone who advocates ontological naturalism about mature moral theory. Michael Smith's believes that the following is analytic: ϕ -ing in C is right if and only if, if we were fully rational, we would desire to ϕ in C. But, his account is non-reductive, since he allows that our account of what it takes to be fully rational need not itself be reductive. He merely claims that the right hand side of his equivalence gives us a summary of all the relevant platitudes surrounding the term 'right'. However, his further claim is that this non-reductive analysis of moral terms can be squared with 'a broader naturalism', since the property of full rationality – the state of having a maximally 'coherent, unified and systematically justified' set of desires, i.e. having reached a mature moral theory – *can* be identified with a natural property. He says that we can 'identify these normative features of a fully rational creature's psychology with natural features of his psychology' (See p. 186). Thus, I suggest, Smith's theory is ultimately a version of ontological naturalism about mature moral theory.

supervenience and neither does even stronger versions of substantive supervenience. The most viable form of analytical naturalism – moral functionalism, the only form of analytical naturalism that survives the open question argument – also proved unable to impose disputational supervenience, since it had seemingly insuperable, independent, problems. Thirdly, ontological naturalism does not give us an *a priori* equivalence and, hence, is a non-starter in the race for disputational supervenience. Finally, the two forms of naturalism were both shown to lead to highly problematic, relativistic, conceptions of morals. In sum, there appears to be no immediate threat to moral minimalism from any of these quarters. That is, there is seemingly no way in which a substantive naturalistic theory of moral discourse threatens to give us an *a priori* guarantee that all moral disagreements will be cognitively blameworthy.

Bibliography

- Altham, J.E.J.: (1986) "The Legacy of Emotivism" in MacDonald and Wright (eds.): *Fact, Science and Morality*, (Oxford: Blackwell, 1986).
- Barcan Marcus, R.: (1980) "Moral Dilemmas and Consistency", *Journal of Philosophy* 77, p. 121-36 and Gowans (1987a).
- Blackburn, Simon:
- (1973) "Moral Realism" in Blackburn (1993).
 - (1984) *Spreading the Word*, (Oxford: Oxford University Press, 1984)
 - (1985) "Supervenience Revisited", in Blackburn (1993).
 - (1988) "How to be an Ethical Antirealist" in Blackburn (1993).
 - (1993) *Essays in Quasi-Realism*, (Oxford: Oxford University Press, 1993)
 - (1994) "Circles, Finks, Smells and Biconditionals", in Tomberlin (ed.) *Philosophical Perspectives*: vol. 7, *Language and Logic*.
 - (1996) "Dilemmas: Dithering, Plumping, and Grief", in Mason (1996).
 - (1997) "Wittgenstein, Wright, Rorty and Minimalism", *Mind*, vol. 107.
 - (1998) "Moral Relativism and Moral Objectivity", *Philosophy and Phenomenological Research* vol. 58, no. 1, March.
 - (1998) *Ruling Passions*, (Oxford: Oxford University Press, 1998).
- Blackburn, Simon and Simmons, Keith (eds.): (1999) *Truth*, (Oxford: Oxford University Press, 1999)
- Boyd, Richard: (1988) "How to be a Moral Realist", in Darwall, Gibbard, Railton (1997).
- Brink, David O.: (1996) "Moral Conflict and Its Structure", in Mason (1996).
- Dancy, Jonathan:
- (1993) *Moral Reasons*, (Oxford: Blackwell, 1993).
 - (1994) "Why There Really is no Such Thing as the Theory of Motivation", *Proceedings of the Aristotelian Society*, Vol. 95, 1994-95, 1-18.
 - (1996) "Real Values in a Humean Context", *Ratio (new series)*, vol. 9, 1996, 171-83.
- Davidson, Donald:
- (1984) *Inquiries into Truth and Interpretation*, (Oxford: Oxford University Press, 1984).
 - (1980a) "Actions, Reasons and Causes", in *Essays on Actions and Events*, (Oxford: Oxford University Press, 1980).

— (1980b) "How is Weakness of the Will Possible?", in *Essays on Actions and Events*, (Oxford: Oxford University Press, 1980).

Divers, John and Miller, Alexander:

— (1994) "Why Expressivist about Value Should Not Love Minimalism about Truth", *Analysis*, vol. 54.1, 1994, 12-19.

— (1995) "Platitudes and Attitudes", *Analysis*, vol. 55.1, 1995, 37-44.

— (1999) "Arithmetical Platonism: Reliability and Judgement-Dependence", *Philosophical Studies*, forthcoming

Darwall, Stephen, Gibbard, Alan and Peter Railton (ed.):

— (1992) "Toward *Fin de Siècle* Ethics: Some Trends", *The Philosophical Review*, vol. 101, 115-189 and in Darwall, Gibbard, Railton (1997).

— (1997) *Moral Discourse and Practice – Some Philosophical Approaches*, (Oxford: Oxford University Press, 1980).

Darwall, Stephen:

— (1998) "Expressivist Relativism?", *Philosophy and Phenomenological Research*, vol. 58.1, March.

— (1998) "The Error-theory and Ethical Relativism", in Darwall: *Philosophical Ethics*, (Westview Press, 1998).

Dreier, James:

— (1990) "Internalism and Speaker Relativism", *Ethics*, vol. 101, 6-26.

— (1996) "Review of Smith: The Moral Problem", *Mind*, vol. 105. April 1996, 363-367.

Dretske, Fred: (2000) "Entitlement: Epistemic Rights Without Epistemic Duties?", *Philosophy and Phenomenological Research*, vol. 60, 3, May, 2000.

Dummett, Michael: (1959) "Truth", in *Truth and other Enigmas*, (London: Duckworth, 1978).

Field, Hartry: (1989) *Realism, Mathematics and Modality*, (Oxford: Blackwell, 1989).

Foot, Philippa:

— (1983) "Moral Realism and Moral Dilemma", *Journal of Philosophy*, vol. 80, p. 379-98 and Gowans (1987a).

— (1995) "Moral Dilemmas revisited", in Sinnott-Armstrong, Raffman and Asher (eds.): *Morality, Modality and Belief*, (Cambridge University Press, 1995).

Gibbard, Alan:

— (1990) *Wise Choices, Apt Feelings*, (Cambridge, Mass.: Harvard University Press, 1990).

- (1996) "Hale and Blackburn on the Geach-Searle Problem", from conference-paper, Warwick 1996.
- Gowans, C. W.:
- (1987a) (ed.) *Moral Dilemmas*, (Oxford: Oxford University Press, 1987).
- (1987b) "The Debate on Moral Dilemmas" in Gowans (1987a).
- Griffin, James: (1996) *Value Judgement. Improving Our Ethical Beliefs*, (Oxford: Oxford University Press, 1996).
- Haldane, John and Wright, Crispin (eds.):
- (1993) *Reality, Representation and Projection*, (Oxford: Oxford University Press, 1993).
- Hale, Bob:
- (1986) "The Compleat Projectivist", *Philosophical Quarterly*, vol. 36.142, 65-84.
- (1987) *Abstract Objects*, (Oxford: Blackwell, 1987).
- Hale, Bob and Wright, Crispin:
- (1997) *A Companion to the Philosophy of Language*, (Oxford: Blackwell, 1997).
- Hare, R. M.:
- (1963) *Freedom and Reason*, (Oxford: Oxford University Press, 1963).
- (1972) "Principles", *Proceedings of the Aristotelian Society*, vol. 73, 1-18.
- (1981) *Moral Thinking*, (Oxford: Oxford University Press, 1981).
- (1984) "Supervenience", *Proceedings of the Aristotelian Society*, supp. vol. 56, 1-16.
- (1999) *Objective Prescriptions and other Essays*, (Oxford: Oxford University Press, 1999).
- Harman, Gilbert:
- (1975) "Moral Relativism Defended", *Philosophical Review*, vol. 84, 3-22.
- (1977) *The Nature of Morality*, (New York: Oxford University Press, 1977).
- (1978) "What is Moral Relativism?" in Goldman and Kim (eds.): *Values and Morals*, (Dordrecht, Netherlands: D. Reidel, 1978).
- (1989) "Is there a Single True Morality?" in Copp and Zimmermann (eds.) *Morality, Reason and Truth*, (Totowa, N.J.: Rowman and Allenheld, 1985).
- (1998) "Precis of Part One" and "Responses to Critics", *Philosophy and Phenomenological Research*, vol. 58. 1.,
- Harman, Gilbert and Judith Jarvis Thompson:
- (1996) *Moral Relativism and Moral Objectivity*, (Oxford: Blackwell, 1996).
- Horgan, Terrence and Timmons, Mark:
- (1992) "Troubles for New Wave Moral Semantics: The 'Open Question Argument' Revived", *Philosophical Papers*, vol. 21, 153-75.

Horwich, Paul:

- (1990) *Truth*, (Oxford: Blackwell, 1990).
- (1994) "The Essence of Expressivism", *Analysis*, vol. 54.1, 1994, 19-20.

Hume, David: (1739-40) *A Treatise of Human Nature*, Book III, (Harmondsworth, Middlesex: Penguin, 1987).

- (1777) *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, (Oxford: Oxford University Press, 1975).

Jackson, Frank:

- (1994) "Realism, Truth and Truth Aptness", *Philosophical Books*, 162-69.
- (1998) *From Metaphysics to Ethics – A Defence of Conceptual Analysis*, (Oxford: Oxford University Press, 1998).

Jackson, Frank, Oppy, Graham and Smith, Michael:

- (1994) "Minimalism and Truthaptness", *Mind*, vol. 103., July 1994, 287-302.

Jackson, Frank and Pettit, Phillip:

- (1988) "Functionalism and Broad Content", *Mind*, vol. 97, no. 387.
- (1995) "Moral Functionalism and Moral Motivation", *Philosophical Quarterly*, vol. 45, 20-40.
- (1996) "Moral Functionalism, Supervenience and Reductionism", *Philosophical Quarterly*, p. 82-86.

Johnston, Mark:

- (1989) "Dispositional Theories of Value", *Proceedings of the Aristotelian Society*, supp. vol. 63, 139-74.
- (1991) "Explanation, Response-dependence and Judgement-dependence", in Menzies (ed.), *Response-dependent Concepts*, ANU, Working Papers in Philosophy, vol. 1.

Kölbel, Max: (1997) "Wright's argument from Neutrality", *Ratio* (New Series), vol. 10, 35-47.

Kripke, Saul: (1980) *Naming and Necessity*, (Cambridge, Mass.: Harvard University Press, 1980).

Lewis, David:

- (1989) "Dispositional Theories of Value", *Proceedings of the Aristotelian Society*, supp. vol. 63, 113-37.
- (1996) "Desire as Belief II", *Mind*, vol. 105, April 1996, 303-313.

Mackie, J.L.: (1977) *Ethics: Inventing Right and Wrong*, (Harmondsworth, Middlesex: Penguin, 1977).

- Mason, H.E. (ed.): (1996) *Moral Dilemmas and Moral Theory*, (Oxford: Oxford University Press, 1996).
- McDowell, John:
- (1978) "Are Moral Requirements Hypothetical Imperatives?", *Proceedings of the Aristotelian Society*, supp. vol. 52, 13-29.
 - (1979) "Virtue and Reason", *The Monist*, vol. 62, 331-50.
 - (1981) "Non-Cognitivism and Rule-Following", in Holtzman and Leich: *Wittgenstein: To Follow a Rule*, (Boston: Routledge & Kegan Paul, 1981), 141-62.
 - (1985) "Values and Secondary Qualities", in T. Honderich (ed.) *Morality and Objectivity*, (London: Routledge & Kegan Paul, 1985).
 - (1986) "Critical Notice of *Ethics and the Limits of Philosophy*", *Mind*, 95, 377-86.
 - (1987) "Projection and Truth in Ethics", The Lindley Lecture, The University of Kansas, and in Darwall, Gibbard, Railton (1997).
 - (1994) *Mind and World*, (London, England: Harvard University Press, 1994).
 - (1998) *Mind, Value and Reality*, (Cambridge, Massachusetts: Harvard University Press, 1998).
- McFetridge, I.G.: (1985) "Supervenience, Realism, Necessity", *The Philosophical Quarterly*, vol. 35, 245-58..
- McNaughton, David: (1988) *Moral Vision*, (Oxford: Blackwell, 1988).
- Miller, Alexander: (1996) "An objection to Smith's Argument for Internalism", *Analysis*, 56.3, 169-74.
- Moore, G.E.: (1903) *Principia Ethica*, (Cambridge, England: Cambridge University Press, 1903).
- Pettit, Phillip:
- (1991a) "Realism and Response-Dependence", *Mind*, vol. 100.4, October,
 - (1991b) "Consequentialism", in Singer (ed.): *A Companion to Ethics*, (Oxford: Blackwell, 1991).
 - (1996) "Realism and Truth: A Comment on Crispin Wright's Truth and Objectivity", *Philosophy and Phenomenological Research* vol. 56.4, December, 883-90.
- O'Leary-Hawthorne and Price, Huw:
- (1996) "How to Stand Up for Non-cognitivists", *Australasian Journal of Philosophy*, vol. 74.2, .
- Parfit, Derek: (1984) *Reasons and Persons*, (Oxford: Oxford University Press, 1984).
- Platts, Mark: (1979) *Ways of Meaning*, (London: Routledge and Kegan Paul, 1979).

Price, Huw:

- (1987) "Truth and the Nature of Assertion", *Mind* vol. 96, 202-20.
- (1988) *Facts and the Function of Truth*, (Oxford: Blackwell, 1988).

Putnam, Hilary:

- (1992a) *Realism with a Human Face*, (Cambridge, Mass.: Harvard University Press, 1992).
- (1992b) *Renewing Philosophy*, (Cambridge, Mass.: Harvard University Press, 1992).

Railton, Peter:

- (1986) "Moral Realism", *Philosophical Review*, vol. 95, 163-207.
- (1993a) "What the Non-Cognitivist Helps Us to See the Naturalist Must Help Us to Explain", in Haldane and Wright (1993).
- (1993b) "Reply to David Wiggins", in Haldane and Wright (1993).
- (1996) "The Diversity of Moral Dilemma", in Mason (1996).
- (1998) "Moral Explanation and Moral Objectivity", *Philosophy and Phenomenological Research*, vol. 58.1, March.

Rawls, John: (1972) *A Theory of Justice*, (Oxford: Oxford University Press, 1972).

Sainsbury, Mark: (1996) 'Crispin Wright: Truth and Objectivity', *Philosophy and Phenomenological Research*, vol. 56.4, December, 899-904.

Sayre-McCord, G.:

- (1986) "The Many Moral Realisms", Spindel Conference: Moral Realism, *Southern Journal of Philosophy*, vol. 24, suppl., 1-22
- (1991) "Being a Realist about Relativism (in Ethics)", *Philosophical Studies*, vol. 61, 155-176.

Scanlon, T. M.: (1998) *What We Owe to Each Other*, (Cambridge, Mass.: Harvard University Press, 1998).

Schafer-Landau, Russ:

- (1994) "Ethical Disagreement, Ethical Objectivism and Moral Indeterminacy", *Philosophy and Phenomenological Research*, 331-46.
- (1995) "Vagueness, Borderline Cases and Moral Realism", *American Philosophical Quarterly*, vol. 32.1,

Shapiro, Stewart and Taschek, Bill:

- (1996) "Intuitionism, Pluralism and Cognitive Command", *Journal of Philosophy*, vol. 93, 74-88.

Sinnott-Armstrong, Walther:

— (1988) *Moral Dilemmas*, (Oxford: Blackwell, 1988).

— (1996) "Moral Dilemmas and Rights", in Mason (1996).

Skorupski, John:

— (1986) "Objectivity and Convergence", *Proceedings of the Aristotelian Society*, vol. 86, 235-50.

— (1997) "Reason and Reasons" in Cullity and Gaut (eds.): *Ethics and Practical Reason*, (Oxford: Oxford University Press, 1997).

— (1999) "Irrealist Cognitivism" *Ratio (new series)*, vol. 12, 4 December, 1999, 436-459.

Smith, Michael:

— (1987) "The Humean Theory of Motivation", *Mind*, 1987, 36-61.

— (1989) "Dispositional Theories of Value", *Proceedings of the Aristotelian Society*, supp. vol. 63, 89-111.

— (1994a) "Why Expressivist about Value Should Love Minimalism about Truth", *Analysis*, vol. 54.1, 1-12.

— (1994b) "Minimalism, Truth-aptness and Belief", *Analysis*, vol. 54.1, 21-26.

— (1994c) *The Moral Problem*, (Oxford: Blackwell, 1994).

— (1995) "Internalism's Wheel", *Ratio (new series)*, vol. 8, December 1995, 277-302.

— (1996a) "The Argument for Internalism: reply to Miller", *Analysis*, vol. 56.3, 175-184.

— (1996b) "Normative Reasons and Full Rationality: Reply to Swanton", *Analysis*, vol. 56.3, 175-184.

Stocker, Michael: (1990) *Plural and Conflicting Values*, (Oxford: Oxford University Press, 1990).

Sturgeon, Nicholas:

— (1985) "Moral Explanations", in Copp and Zimmerman (eds.) *Morality, Reason and Truth*, (Totowa N.J.: Rowman and Allenheld, 1985).

— (1994) "Moral Disagreement and Moral Relativism", *Social Philosophy and Policy*, vol. 11, 80-115.

— (1998) "Thompson Against Moral Explanations", *Philosophy and Phenomenological Research*, vol. 58.1, March, .

Tennant, Neil: (1997) *The Taming of the True*, (Oxford: Oxford University Press, 1997).

Tersman, Folke: (1998) "Crispin Wright on Moral Disagreement", *Philosophical Quarterly*, vol. 48, 359-65.

- Thomas, Alan: (1997) "Minimalism versus Quasi-realism: Why the Minimalist has a Dialectical Advantage", *Philosophical Papers*, vol. 26.3, .
- Thompson, J.J.: (1998) "Precis of Part Two" and "Reply to Critics", *Philosophy and Phenomenological Research*, vol. 58.1, March, .
- Timmons, Mark: (1999) *Morality without Foundations*, (New York: Oxford University Press, 1999).
- Van Fraassen, Bas: (1973) "Values and the Heart's Command", *Journal of Philosophy*, vol. 70, 5-19 and Gowans (1987a).
- Wiggins, David:
- (1976) "Truth, Invention and the Meaning of Life", *Proceedings of the British Academy*, vol. 62. Also in Wiggins (1998c)..
 - (1990) "Moral Cognitivism, Moral Relativism and Motivating Moral Beliefs", *Proceedings of the Aristotelian Society*, vol. 91, 1990-91, 61-85.
 - (1993a) "Cognitivism, Naturalism, and Normativity: A Reply to Peter Railton", in Haldane and Wright (1993).
 - (1993b) "A Neglected Position?", in Haldane and Wright (1993).
 - (1995) "Objective and Subjective in Ethics", *Ratio (new series)*, vol. 8, December 1995, 243-58.
 - (1998a) "Truth, and Truth as Predicated of Moral Judgements", in Wiggins (1998c).
 - (1998b) "A Sensible Subjectivism?", in Wiggins (1998c).
 - (1998c) *Needs, Values, Truth*, (Third edition), (Oxford: Oxford University Press, 1998).
- Williams, Bernard:
- (1965) "Ethical Consistency", *Proceedings of the Aristotelian Society*, supp. vol. 39, 103-24 and in Gowans (1987a) and Williams (1973).
 - (1966) "Consistency and Realism", *Proceedings of the Aristotelian Society*, supp. vol. 40, 1-22, and in Williams (1973).
 - (1972) *Morality: An Introduction to Ethics*, (New York: Cambridge University Press, 1972).
 - (1973) *Problems of the Self*, (Cambridge: Cambridge University Press, 1973)
 - (1981) *Moral Luck*, (New York: Cambridge University Press, 1981).
 - (1985) *Ethics and the Limits of Philosophy*, (Cambridge, Mass.: Harvard University Press, 1985).
 - (1995a) *Making Sense of Humanity*, (Cambridge: Cambridge University Press, 1995).

— (1995b) "Truth in Ethics", *Ratio (new series)*, vol. 8, December 1995, 227-242.

Williamson, Timothy:

— (1994a) "Critical Study of Truth and Objectivity", *International Journal of Philosophical Studies*, vol. 30.1, 130-44.

— (1994b) *Vagueness*, (London: Routledge, 1994).

— (1996) "Unreflective Realism", *Philosophy and Phenomenological Research*, vol. 56.4, December, 905-909.

Wittgenstein, Ludwig: *Philosophical Investigations*, (Oxford: Blackwell, 1958).

Wong, David: (1991) "Commentary on Sayre-McCord: "Being a Realist about Relativism", *Philosophical Studies*, vol. 61, 177-166.

Wright, Crispin:

— (1988) "Moral Values, Projection and Secondary Qualities", *Proceedings of the Aristotelian Society*, supp. vol. 62. 1-26.

— (1992) *Truth and Objectivity*, (Cambridge, Mass.: Harvard University Press, 1992).

— (1993) *Realism, Meaning and Truth*, (Second edition), (Oxford: Blackwell, 1993).

— (1994a) "Realism, Pure and Simple? A Reply to Williamson", *International Journal of Philosophical Studies*, vol. 30.2, 147-61.

— (1994b) "The Epistemic Conception of Vagueness", *Southern Journal of Philosophy*, vol. 33, suppl.

— (1994c) "Response to Jackson", *Philosophical Books*, 169-75.

— (1995) "Truth in Ethics", *Ratio (New Series)*, vol.8, December 1995, 209-26.

— (1996a) "Précis of Truth and Objectivity", *Philosophy and Phenomenological Research*, vol. 56.4, December, 863-68.

— (1996b) "Judgement-dependence and Anomaly", forthcoming.

— (1996c) "Response to Commentators", *Philosophy and Phenomenological Research*, vol. 56.4, December, 911-41.

— (1996) "Review Article: Human Nature?", *The European Journal of Philosophy*, vol.4, 235-54.

— (1998a) "Comrades against Quietism", *Mind*, vol. 107, 183-203.

— (1998b) "Truth: a Traditional Debate Reviewed", in Blackburn and Simmons (1999).