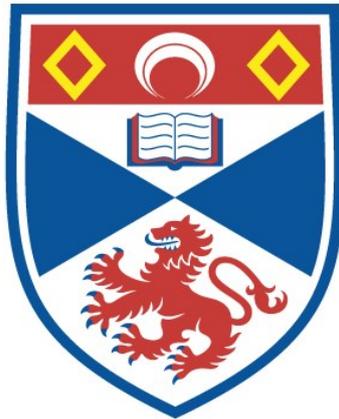


**CONTEMPORARY ETHICAL NATURALISM : A
COMPARATIVE METAETHICAL EVALUATION OF
EXPRESSIVISM AND CORNELL REALISM**

Jorn Sonderholm

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



2005

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

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

Please use this identifier to cite or link to this item:

<http://hdl.handle.net/10023/13224>

This item is protected by original copyright

CONTEMPORARY ETHICAL NATURALISM

A Comparative Metaethical Evaluation of Expressivism and Cornell Realism

Jorn Sonderholm

Thesis for the Ph.D. Degree in Philosophy
School of Philosophical and Anthropological Studies
University of St. Andrews

August 8, 2004



ProQuest Number: 10166938

All rights reserved

INFORMATION TO ALL USERS

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

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



ProQuest 10166938

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

All rights reserved.

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

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

Th
E762

Declarations

I¹
6/000 hereby certify that this thesis, which is approximately
6/000 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. A A

date 6/8-04 signature of candidate ..

I was admitted as a research student in 09.00 [month, year] and as a
candidate for the degree of ... Ph.D. in 09.01 [month, year]; the
higher study for which this is a record was carried out in the University of St
Andrews between 2000.... [year] and 2004.... [year]. A A

date 6/8-04 signature of candidate

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

date 5/8/04 signature of supervisor //

- (d) The copyright of a thesis normally belongs to its author, but candidates must attach to each copy of the thesis a signed declaration in one of the forms which follow (A is normal; B may only be used after consultation with your supervisor and the Head of School and with the permission of the Faculty):

A Unrestricted

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. A A A

date 6/8-04 signature of candidate .

Handwritten mark

Abstract

This thesis contains a critical discussion of two metaethical theories: expressivism, as developed in the works of Simon Blackburn, and Cornell realism, as presented by Richard Boyd and David Brink. In the introduction, a distinction is made between external and internal accommodation projects for moral discourse and it is argued that the external accommodation project should be guided by acceptance of methodological naturalism. Expressivism and Cornell realism are then subjected to an extended comparative evaluation, and an answer is sought to the question of which of the two should be favoured. The main conclusion of the thesis is that Cornell realism is rationally preferable to expressivism. This conclusion is arrived at by looking at how well the two theories, respectively, explain various deeply embedded features of moral discourse. Explaining such features is what the internal accommodation project for moral discourse consists in.

The assertoric surface-form of moral discourse and the supervenience of moral predicates on natural predicates receive special attention in the study. It is argued that expressivism and Cornell realism do equally well on the issue of moral supervenience. But whereas expressivism is still vulnerable to a particular argument from the philosophy of language (the Frege-Geach point), Cornell realism can fend off the criticism that most persistently has been directed at it from this area of philosophy. In a comparative evaluation involving the selected issues, Cornell realism therefore fares better than expressivism.

Acknowledgements

I am indebted to Sarah Broadie, Garrett Cullity, Bob Hale, Daniel Nolan, Agustin Rayo, John Skorupski and Crispin Wright each of whom offered valuable criticism of drafts of individual chapters. I owe a special thanks to my supervisor John Haldane. Throughout my time in St Andrews, he has read and commented on my work. He also carefully commented on a draft of the whole thesis. I have learned a lot from him.

Contents

Acknowledgements

Chapter 1:	Introduction	
	1. <i>Setting the Stage</i>	1
	2. <i>The Problem</i>	4
	3. <i>Thesis Outline</i>	6
Chapter 2:	Naturalism	
	1. <i>Introductory</i>	8
	2. <i>Two Types of Naturalism</i>	8
	3. <i>A Characterisation of Ontological Naturalism</i>	11
	4. <i>The Discipline Criterion</i>	13
	5. <i>The Causal Criterion</i>	24
	6. <i>Methodological Naturalism</i>	26
Chapter 3:	Supervenience	
	1. <i>Introductory</i>	32
	2. <i>An Initial Characterisation</i>	32
	3. <i>Types of Supervenience</i>	36
	4. <i>Ascriptive Supervenience</i>	45
	5. <i>A Formal Feature of Weak Supervenience</i>	47
	6. <i>Supervenience and Ethics</i>	52
Chapter 4:	Realism, Expressivism and Moral Supervenience	
	1. <i>Introductory</i>	56
	2. <i>Blackburn's Argument</i>	57
	3. <i>Responses to the Argument</i>	61
	4. <i>Expressivism</i>	66
	5. <i>The Expressivist Explanation of Moral Supervenience</i>	71

Chapter 5:	Round One of the Frege-Geach Point	
	1. <i>Introductory</i>	76
	2. <i>The Problem</i>	77
	3. <i>The Formal Language E_{ex}</i>	81
	4. <i>Critiques of Blackburn's Response</i>	85
	5. <i>An Alternative Analysis of 1)</i>	87
Chapter 6:	Round Two of the Frege-Geach Point	
	1. <i>Introductory</i>	91
	2. <i>A New Formal Language</i>	92
	3. <i>Features of the Language</i>	94
	4. <i>Validity, Consistency and the Merits of the Language</i>	101
	5. <i>A New Semantics For Connectives</i>	105
	6. <i>A Problem With Rule 4)</i>	108
	7. <i>Further Critiques of the Formal Language</i>	112
	8. <i>Concluding Remarks</i>	119
Chapter 7:	Round Three of the Frege-Geach Point	
	1. <i>Introductory</i>	120
	2. <i>First Proposal</i>	121
	3. <i>Second Proposal</i>	123
	4. <i>A Reconstruction of the Argument</i>	125
	5. <i>Blackburn's Response</i>	127
	6. <i>A Problem for Blackburn's Response</i>	130
	7. <i>Expressivism Entails Revisionism</i>	134
Chapter 8:	Ethical Naturalism and the Open Question Argument	
	1. <i>Introductory</i>	136
	2. <i>Realism</i>	136
	3. <i>Narrow Reduction</i>	140
	4. <i>The Open Question Argument</i>	142
	5. <i>Another Alternative</i>	147

Chapter 9:	The Semantic Views of Cornell Realism	
	1. <i>Introductory</i>	149
	2. <i>A Causal-Functional Account of Moral Terms</i>	150
	3. <i>Putnam's Twin-Earth Example</i>	153
	4. <i>Timmons' Argument</i>	156
	5. <i>Replies to Timmons' Argument</i>	159
Chapter 10:	Cornell Realism and Moral Supervenience	
	1. <i>Introductory</i>	165
	2. <i>Moral Properties as Natural Properties</i>	166
	3. <i>Weak and Strong Supervenience</i>	173
	4. <i>Cornell Realism and Moral Supervenience</i>	178
	5. <i>The Real Explanatory Challenge</i>	180
	6. <i>Brink's Response</i>	183
	7. <i>Final Remarks</i>	189
Chapter 11:	Conclusion	
	1. <i>Introductory</i>	190
	2. <i>Summary and Main Conclusions</i>	190
	3. <i>Resisting the Conclusions and a Short Glance Ahead</i>	192
	4. <i>Metaethics and Normative Ethics</i>	196
Bibliography		203

CHAPTER 1

Introduction

1. *Setting the Stage*

This thesis concerns metaethics. It raises and discusses theoretical issues concerning the nature of moral discourse, and does not offer any guidance in relation to issues in moral theory or applied ethics. The conclusions of the thesis are therefore of little, if any, direct relevance to our everyday, practical life as moral agents. First-order reflections about what to do, or about how to live, do not, however, stand alone, or even particularly prominent, in relation to the question of what is interesting and important from a philosophical point of view. There is a distinct philosophical task that consists in taking a step back from our first-order reflections, and ask second-order questions about the nature of these reflections. Such second-order questions include, but are not restricted to, ones like these: can moral utterances be evaluated in terms of truth and falsity, and if so, what notion of truth is appropriate to moral discourse? Does a notion of moral truth require for its intelligibility, the existence of (perhaps, distinctively moral) states of affairs such that moral claims are true just in case they give a correct depiction of these states of affairs? What is the relation between moral language and reality? Do moral utterances describe states of affairs, or are they used to express emotions or something else with non-propositional content? Can we refer to the classical notions of inconsistency and entailment when we are to explain the validity of certain patterns of inference used in

moral reasoning, or do we have to say that a non-classical system of logic applies to inferences in this area of discourse?

What distinguishes one metaethical theory from another is the nature of the answers it offers to questions like these. In the following, I shall be concerned with assessing various answers to some of these questions, and I shall suggest that some answers are better than others. A major assumption of the thesis is that a view I call 'methodological naturalism' should be accepted. Acceptance of methodological naturalism sets constraints on what views one can hold in areas of thought such as philosophy of mathematics, philosophy of religion, aesthetics and metaethics. If we focus on metaethics, it seems to me that an adequate metaethical theory must meet two desiderata. First of all, it must be able to accommodate features of ordinary moral discourse and practice that are commonly taken to be characteristic of the discourse. Such features include the assertoric surface-form of moral discourse, the existence of the truth-predicate in moral discourse, and the phenomenon that it is impossible that two objects should differ in terms of moral properties without differing in terms of natural properties. A metaethical theory that is unable to explain features such as these, yields a distorted picture of moral discourse and practice. This first desideratum can be summed up in the following way:

"A plausible metaethical view should comport with deeply embedded presumptions of ordinary moral discourse and practice. This guides the project of internal accommodation" (Timmons 1999:12).

The second desideratum that an adequate metaethical theory must meet is that it must not conflict with well-supported non-moral theories and assumptions. The idea is that a metaethical theory that is based on implausible epistemological and ontological

assumptions, cannot be satisfactory. This second desideratum might be formulated like this:

"A plausible metaethical view should comport with plausible general views and assumptions from other relevant areas of inquiry. This guides the project of external accommodation" (ibid.:12).

If methodological naturalism is accepted, as I argue that it should be, then it is the truth of methodological naturalism that should guide the project of external accommodation. That methodological naturalism guides the project of external accommodation means in particular that it is a requirement on an acceptable metaethical theory that it only involves methods of acquiring knowledge that are ones employed by the best contemporary empirical science, and that it only posits the existence of entities that are referred to in the explanatory theories of the best contemporary empirical science.¹

Working within the constraints of methodological naturalism constitutes a kind of 'play safe' strategy. It is to say: 'certain methods of acquiring knowledge are well tested and have proven very reliable. We also know that things such as atoms, acids, desires, beliefs, biological needs, behavioural patterns and societies exist. Let us see how many of the characteristic features of ordinary moral discourse we can explain and make sense of if we only make use of these tested methods of acquiring knowledge, and if we only posit the existence of entities that we know are there'. It might of course be the case that in the end it proves impossible to explain and make sense of the characteristic features of ordinary moral discourse, if we work within the boundaries of methodological naturalism. If this turns out to be the case, then we are faced with the choice of either abandoning methodological naturalism or abandoning

¹ This view is spelled out and defended in the next chapter.

moral discourse. This choice only becomes pressing, however, when the failure has been established of every attempt to carry out the project of internal accommodation, constrained by the core claims of methodological naturalism.

2. The Problem

Historically, many attempts have been made to carry out the external accommodation project within a naturalistic framework. On the contemporary metaethical scene, Simon Blackburn and the philosophers associated with the school of thought that has come to be known as 'Cornell realism', are examples of theorists who are engaged in this project.² For a number of years, Blackburn has defended a theory labelled 'quasi-realism'. The Cornell realists defend a version of what is called 'synthetic naturalism'. Blackburn and the Cornell realists offer their own characterisations of what naturalism is, but it is safe to say that neither Blackburn's theory nor Cornell realism transgress the boundaries of an external accommodation project which is guided by acceptance of methodological naturalism. Put slightly differently: both Blackburn's theory and Cornell realism can be seen as attempts to carry out the external accommodation project when this project is guided by the truth of methodological naturalism. Consequently both meet one of the desiderata that any minimally acceptable metaethical theory should meet.

The main question of the thesis can now be formulated: given that one is attracted to methodological naturalism, and has a choice in metaethics between Blackburn's quasi-realism and Cornell realism, which of these two theories should be favoured? This question is to be decided by looking at how well the two theories, respectively, carry out the internal accommodation project. Given that both theories

² Cornell realism is commonly associated with the work of David Brink, Richard Boyd, Nicholas Sturgeon and Richard Millar.

meet the desideratum of being compatible with methodological naturalism, the question of which theory to favour must be answered by comparing how well the two theories, respectively, explain and account for deeply embedded presumptions of ordinary moral discourse and practice.

Perhaps some justification is needed for this framework. There are, after all, other metaethical theories on the contemporary scene that meet the desideratum of being compatible with methodological naturalism.³ Why not discuss these? My reasons for focusing on quasi-realism and Cornell realism are these: non-descriptivism, which is a core component of Blackburn's quasi-realism, has been a prominent line of thought in the history of philosophy. Blackburn can be seen to carry the flag, previously carried by Hume, Ayer, Stevenson and Hare. Blackburn's writings have, moreover, been very influential over the past two decades and Blackburn continues to be an important figure. When it comes to Cornell realism, it can also be said that this is a version of a general line of thought that has been prominent in the history of philosophy. Realism and descriptivism in ethics have been around since, at least, Plato, and it is proper and interesting to contrast a contemporary anti-realist and non-descriptivist position with a contemporary realist and descriptivist position. Brink and Boyd, who are the two theorists I shall focus on in my discussion of Cornell realism, continue, moreover, to be influential writers in the field.⁴ There is therefore good reason to choose Cornell realism as one of the positions to be discussed.

³ See, for example, (Railton 1986; 1989), (Gibbard 1990), (Jackson 1998), (Smith 1994), (Skorupski 1999), (Timmons 1999) and (Wright 1996).

⁴ See (Brink 2001) and (Boyd 2003).

3. Thesis Outline

Having made clear what the main question of the thesis is, and having offered some justification for the criteria by which expressivism and Cornell realism have been chosen for examination and comparison, it is now possible to turn to an outline of how the thesis is structured in terms of individual chapters. The next chapter contains a discussion of what naturalism is as a general ontological and methodological view. The key features of methodological naturalism, as I shall understand the position, are specified, and the position is defended. In chapter three, I start focusing on the project of internal accommodation. There are a number of deeply embedded presumptions of ordinary moral discourse and practice that a plausible metaethical theory should comport with. One such presumption revolves around the concept of supervenience. A general introduction to this concept is the topic of this chapter. A particular definition of moral supervenience is suggested, and it is argued that any minimally adequate metaethical theory must be able to explain why this relation holds. Chapter four contains a discussion of Simon Blackburn's influential supervenience argument against moral realism. It also contains an account of what expressivism is as a semantic theory, and a discussion of the expressivist explanation of moral supervenience. I argue that Blackburn's argument is powerful, and that expressivism does not have an independent problem on the issue of moral supervenience.

In chapters five, six and seven, attention is directed at the commonly termed 'Frege-Geach point'. I give an account of what this point is, and why it constitutes a challenge to expressivism. I consider thereafter, in some technical detail, the various argumentative strategies that Blackburn, over the years, has employed in order to overcome the challenge. In chapter seven, I discuss Blackburn's most recent strategy.

The conclusion of that discussion is that this strategy does not enable the expressivist to overcome the Frege-Geach point. This is one of the main conclusions of the thesis. At present there simply is no credible solution on the table as to how the expressivist can meet the Frege-Geach point. Expressivism should therefore not be embraced.

In chapter eight, I begin my exploration of the second alternative for the methodological naturalist. I discuss the 'open-question argument', and conclude that this much heralded challenge poses no threat to either analytic ethical naturalism, or to Cornell realism. In chapter nine, I outline and discuss a very influential argument against the plausibility of the semantic views associated with Cornell realism. I argue that this argument, at best, poses an explanatory challenge to Cornell realism, and that this challenge can be met. In addition to an account of some of the key metaphysical commitments of Cornell realism, chapter ten involves a discussion of David Brink's explanation of moral supervenience. The conclusion of this chapter is that the special modal force with which moral supervenience holds, does not amount to a problem for Cornell realism. This position fares well on the issue of moral supervenience. Chapter eleven contains a short summary of the thesis and its main conclusions. The main conclusion of the thesis is that Cornell realism is to be preferred to expressivism as developed by Blackburn. In a comparative evaluation of the two theories, involving the issues I have chosen for discussion, Cornell realism comes out the strongest.

CHAPTER 2

Naturalism

1. *Introductory*

In this chapter, I present and defend a certain view about what non-moral assumptions should guide the external accommodation project for moral discourse. The main claim of the chapter is that this project should be guided by an acceptance of naturalism. This claim needs, however, to be qualified in a number of ways, and the rest of the chapter can be seen as an attempt to provide such a qualification. In section two, an important distinction is drawn between two types of naturalism. In sections three, four and five, various popular characterisations of ontological naturalism are discussed. I argue that these characterisations are either uninformative, internally inconsistent, or vague. In section six, I turn to methodological naturalism. I present and reject a particular characterisation of this type of naturalism. After that, I make clear what I mean by 'methodological naturalism', and I then argue that this position should be accepted.

2. *Two Types of Naturalism*

Naturalism is a significant stance in contemporary philosophy. In areas such as philosophy of mind, philosophy of mathematics and metaethics, it is popular to pledge allegiance to naturalism.⁵ But what is naturalism as a general philosophical outlook? Since so many people call themselves 'naturalist' it would be helpful to know what this label entails in terms of substantial commitments. According to *The*

⁵ See, for example, (Papineau 1993), (Brink 2001), (Jackson 1998) and (Blackburn 1998).

Cambridge Dictionary of Philosophy, which aims to capture contemporary analytical usage, naturalism is a twofold view that has a methodological/epistemological component and an ontological component.⁶ The methodological/epistemological component of naturalism is the claim that:

"acceptable methods of justification and explanation are commensurable, in some sense, with those in natural science".

The ontological component of naturalism is the claim that:

"everything is composed of natural entities - those studied in the sciences (on some versions, the natural sciences) - whose properties determine all the properties of things, persons included, abstracta (abstract entities) like possibilia (possibilities) and mathematical objects, if they exist, being constructed of such abstracta as the sciences allow;"

What is the conceptual relationship between methodological and ontological naturalism if the two components are given this characterisation? It is conceptually possible, and indeed very common for the two views to be combined. For example, if one adheres to the methodological component of naturalism, then this is likely to set constraints on what ontological theses one is willing to accept. A methodological naturalist might reject the existence of abstract objects because the only methods of justification and explanation he allows, do not register the existence of abstract objects. In a case like this, a methodological commitment informs ontological commitments. The reverse relation can also hold. One can start with the assumption that everything is composed of the entities that are studied in the natural sciences, and then go on to form the methodological view that the only acceptable methods of

⁶ Cambridge: Cambridge University Press 1995.

justification and explanation are ones that are commensurable, in some sense, with those in natural science.

However, it is conceptually possible that methodological and ontological naturalism should come apart. Consider the, be it unlikely, case of an agent who endorses ontological, but not methodological naturalism. Such an agent thinks that there are other, and acceptable, ways of justification and explanation of the subject matter of natural science, than the ones that are commensurable with those in natural science itself. As an example of such an agent, consider a materialist who, for some reason, thinks that claims about the material world, put forward on the basis of the laying of tarot cards, are methodologically acceptable. Such an agent might be thought to hold an odd combination of views, but that combination of views is not excluded by the characterisations of methodological and ontological naturalism.

As an example of the coherence of being a methodological, but not an ontological naturalist, consider an agent who accepts the existence of a Deity outside space and time. It is conceptually coherent for such an agent to endorse the view that respectable methods of justification and explanation are commensurable, in some sense, with those in natural science. Methodological naturalism, in the above characterisation, does not say that every aspect of reality can, even in principle, be explained by methods commensurable, in some sense, with those in natural science. Accordingly, the truth of methodological naturalism is conceptually compatible with there being aspects of reality that cannot be justified and explained by methods commensurable, in some sense, with those in natural science.

3. A Characterisation of Ontological Naturalism

To shed further light on the distinction between methodological and ontological naturalism, consider this characterisation of what it is to be a naturalist:

"To be a naturalist is to see human beings as frail complexes of perishable tissue, and so part of the natural order. It is therefore to refuse unexplained appeals to mind or spirit, and unexplained appeals to knowledge of a Platonic order of Forms or Norms; it is above all to refuse any appeal to a supernatural order" (Blackburn 1998:49).

Blackburn's comments serve to make clear what the ontological commitments of a naturalist are, and can therefore be seen to shed light on the ontological component of naturalism. The passage does not, however, touch upon issues relating to methodological naturalism, and acceptance of ontological naturalism, in Blackburn's characterisation, is compatible both with acceptance and with rejection of methodological naturalism, as characterised above.

Interestingly, Blackburn's characterisation is mainly negative. He explicates what it is to be an ontological naturalist by giving a list of things that such a naturalist does not believe in. This negative definition is not without merits. It tells us, for example, that Platonism and theism are rejected by the ontological naturalist, and it therefore gives us some understanding of what ontological naturalism is. More needs to be said, however, if we are to have a satisfying characterisation of what ontological naturalism is, and I do not think that such a characterisation can be of a mainly negative nature. One reason why such a characterisation is insufficient is that it leaves unexplained the rejection of various entities. If ontological naturalism is characterised negatively as the view that entities x, y and z do not exist, then there is nothing in the characterisation alone that unifies the rejection of these entities. And if

there is nothing that unifies the rejection of these entities, then it is difficult to see how ontological naturalism can be a systematic view. If x, y and z are rejected on the ground that they all possess a certain property, and if this is made clear in the characterisation of ontological naturalism, then such a position would go some way towards being a systematic view. A characterisation of ontological naturalism that contains a clause that yields a general criteria for what it takes to be a natural entity is therefore preferable to a characterisation that has informative content primarily in virtue of stipulating that certain entities are not tolerated by ontological naturalism.

Another dissatisfying feature of Blackburn's characterisation is that it is rather uninformative in virtue of referring to the notions of the 'natural order' and a 'supernatural order'. Saying that a naturalist sees human beings as part of the natural order, and that he refuses appeal to a supernatural order is only informative on the assumption that it is clear what 'natural' means in these contexts. It is, however, not clear what 'natural' means in these contexts. To use 'natural' in a characterisation of what it is to be a naturalist, without making explicit what this term means, has no explanatory power. Such a manoeuvre just shifts the question from what it is to be a naturalist to what it is to be natural.

A third problem is this: what new information about what it is to be a naturalist, new information in the sense of not already being inherent in the notion of being a naturalist, is actually conveyed by saying that a naturalist refuses any appeal to a supernatural order? Saying this, only yields new information about what it is to be a naturalist on the assumption that one could be naturalist and appeal to something supernatural. Such an assumption is, however, conceptually incoherent. Whatever it exactly is to be naturalist, it is a conceptual truth about what it is to be a naturalist that a naturalist does not appeal to anything supernatural. The moment a naturalist

appeals to something supernatural in his explanations, he simply stops being a naturalist. So, Blackburn's emphasis that a naturalist, "above all" refuses any appeal to a supernatural order, does not cast any new light on what it is to be a naturalist.

4. *The Discipline Criterion*

If Blackburn's characterisation of the ontological component of naturalism is dissatisfying, are there any other, more satisfying ones? One way of explaining what naturalism is, consists in making use of what has been called the 'discipline criterion' (Crisp 1996:115). According to this criterion, naturalism is the view that the only entities that exist are natural entities. Natural entities are then characterised as those entities that are studied and described by natural science. Mark Timmons is an example of a philosopher who uses the discipline criterion to characterise what naturalism is:

"The vague, pre-theoretic idea that the philosophical naturalist tries to articulate and defend is that everything - including any particulars, events, facts, properties, and so on - is part of the natural, physical world that science investigates" (Timmons 1999:12).

Timmons gives a further elucidation of what he takes philosophical naturalism to be:

"The naturalist begins with an ontological presumption about what sorts of particulars, entities, and so forth are fundamental, which can be roughly captured in the following thesis of ontic primacy: The ontologically primary or fundamental entities (properties, facts, etc.) in the world are all part of the subject matter of science" (ibid.:13).

One thing to note about Timmons' initial characterisation of the philosophical naturalist is that it contains reference to the notion of the 'natural' world. The natural

world is seen to be equivalent to the physical world. Timmons' characterisation is therefore informative. Naturalism is equivalent to physicalism. A philosophical naturalist defends the idea that everything is part of the physical world, and as a result of this, a naturalist cannot accept the existence of non-physical entities such as, say, a Cartesian ego. A naturalist must, simply in virtue of being a naturalist, reject the existence of such an entity.

This does not, however, square well with the subsequent thesis of ontic primacy. Since this thesis only says that the primary or fundamental entities in the world are the subject matter of natural science, the thesis allows the existence of entities that are not the subject matter of the natural sciences. Non-physical entities are examples of entities that are not the subject matter of natural science. The thesis of ontic primacy therefore allows the existence of non-physical entities. This means that if what it is to be a naturalist is characterised in terms of subscription to the thesis of ontic primacy, as Timmons suggests it should be, then a naturalist may be able to accept the existence of a non-physical entity such as a Cartesian ego, assuming that it is not a primary or fundamental entity. It is of course possible for such a naturalist to reject the existence of non-physical entities, but if the naturalist does this, then this is not something that can be justified by reference to his naturalist position. When characterised in terms of the thesis of ontic primacy, naturalism does not entail physicalism, though physicalism is compatible with naturalism.

The upshot of this is that whereas Timmons' two characterisations, when considered individually, cast some light on the question of what it is to be a naturalist, they do not, when seen in conjunction, present naturalism as a consistent position. One remedy to this problem is to go for just one of the characterisations, and leave the other out of the picture. Each of the two characterisations suffer,

however, from a flaw that is common to many attempts to characterise what naturalism is. Both of Timmons' characterisations of what it is to be a naturalist involve reference to a notion of the natural sciences. This notion is, moreover, clearly supposed to do a lot of explanatory work in the characterisations. In relation to this, it is important to be aware that an attempt to characterise natural entities in terms of what is the subject matter of the natural sciences can only be informative if it is clear what sciences are natural sciences. But this is not at all clear. The concept of a natural science is vague. To give any of the following, distinct definitions of a natural science is to demonstrate competence in the use of the concept.

Firstly, a natural science can be defined as a science that gets its results from conducting empirical research and carrying out empirical experiments. This definition has as a consequence that sciences such as physics, chemistry, biology and geology are within the boundaries of natural science. But it also has the consequence that sciences such as social psychology and anthropology are natural sciences.

Secondly, a natural science can be defined as a science that studies physical objects. It is a consequence of this definition that social psychology and anthropology cannot be natural sciences. After all, it does seem to be stretching things to say that the emotional and behavioural characteristics of a group, and the social norms of a culture, are physical objects.

Thirdly, a natural science can be defined as a science that studies natural phenomena where natural phenomena are defined as phenomena that lie within the spatio-temporal world. The important contrast is here between the natural and the supernatural where the latter is some, or all, of the phenomena that lie outside the spatio-temporal world. On this definition of a natural science, the class of natural sciences is greatly expanded compared to the previous definitions. All of the social

sciences study phenomena that lie within the spatio-temporal world and are therefore to be seen as natural sciences. The same is true of medical science and many of the disciplines within the humanities. History, literature, cultural studies and art-history are all areas of inquiry that study phenomena that are natural in the sense that they lie within the spatio-temporal world.⁷

An uncontroversial conclusion to draw from these considerations is that reference to an unqualified notion of the natural sciences does relatively little to make it clear what entities are natural, and as a result of this, it does not cast much light on the question of what naturalism is. Timmons just uses an unqualified notion of the natural sciences, and his characterisations of what it is to be a naturalist are therefore vague.⁸ This can be illustrated by taking yet another look at the thesis of ontic primacy. Acceptance of this thesis leads to two very different world views depending upon which definition of natural science is accepted. If the first definition is accepted, then naturalism is a position that allows the view that non-physical entities such as the emotional and behavioural characteristics of a group and the social norms of a culture are ontologically primary or fundamental entities. If the second definition is accepted, then naturalism is a position that does not allow this view. This means that a characterisation of what it is to be a naturalist, through use of the thesis of ontic primacy, only has real informative content if it is put forward in conjunction with a qualification of what a natural science is.

⁷ This definition of what a natural science is leaves it unclear whether philosophy is a natural science. After all, many philosophers think that philosophy's subject matter is not spatio-temporal entities, but rather general, abstract structures.

⁸ To see that Timmons is not alone in making use of an unqualified notion of natural science in an attempt to characterise what entities are natural, see (Moore 1993:92), and consider this passage from Brink: "First, we must distinguish within the class of nonmoral facts and properties between natural and supernatural properties (see Moore 1993:chapters:1,2,4). This distinction is supposed to be one of which we have an intuitive grasp. Natural facts and properties are presumably something like those facts and properties as picked out by the natural and social sciences (broadly conceived);" (Brink 1989:22).

If naturalism is characterised through use of the discipline criterion, then naturalism actually comes in at least three different versions. The first version says that what is natural is what is the actual, historical subject matter of the natural sciences. One might reject this view because one feels that it places unwarranted weight on the contingent fact of what, as it happens, has been of interest to natural scientist (Crisp1996:116). The worry can perhaps be spelled out in more detail like this: if the set of natural entities is the set of entities that actually has been studied by natural science, then it is the case that if natural scientists had chosen to study other things than the things that they actually chose to study, then we would now have a different set of natural entities from the set that we actually have, and this is unsatisfactory.

It seems to me that this line of thought rests on the assumption that the notion of natural entities is only useful if the set of natural entities is a set whose boundaries can be determined once and for all, and by a method that is completely independent of contingent human interests. This assumption is, in my opinion, mistaken. A defender of this first version of the discipline criterion should say that the natural sciences, whatever the natural sciences exactly are, are engaged in the process of acquiring knowledge of reality. This is an ongoing process and progressively, more and more aspects and layers of reality, are laid bare. The set of type of particles and type of forces, governing the behaviour of these particles, that the natural sciences, at any given point in time, have been able to detect and characterise, is the set that constitutes the set of natural entities. This view entails that the set of natural entities changes along with changes in the investigatory powers of the natural sciences. This does not mean, however, that the notion of natural entities is not a useful notion. The set of natural entities is a set of entities that we, at some given point in time have

strong reason think is there, and independently of the fact that the set of natural entities might be revised in light of future discoveries, the set of entities, as it looks like at that time, should be taken into account when views about the nature of reality are then being formulated.

A second version of naturalism emerges if what is natural is what is potentially the subject matter of natural science. Perhaps this way of characterising naturalism should be rejected because "it begs the very question to which we are seeking an answer" (Crisp 1996:116). I think that this is the right response. Consider the question of what is the potential subject matter of natural science. If something 'non-natural', whatever exactly that is, is a potential subject matter of natural science, whatever exactly that is, then natural science can, at least potentially, study and describe something non-natural. But if natural science can do this, then it is difficult to understand why this kind of science should be called 'natural' science. On any plausible view about what natural science is, it must be stipulated that natural science only studies and describes 'nature' or 'the natural'. Whatever 'nature' or 'the natural' exactly is, it is, per definition, something different from 'the non-natural'. So, it is not a plausible suggestion that something non-natural is the potential subject matter of natural science. It is only the natural that is the potential subject matter of natural science. This means that saying that what is natural, is what is the potential subject matter of natural science, is like saying that what is natural is what is natural. This does not cast any light on the question of what it is to be natural, and it therefore casts no light on what naturalism entails in terms of ontological commitments.

What is involved in the third version of naturalism can be illuminated by considering a recent characterisation of naturalism:

"Ontological naturalism takes various forms. We will understand such naturalism in terms of this core view: every real entity either consists of or is somehow ontologically grounded in the objects countenanced by the hypothetically completed empirical sciences" (Moser & Yandell 2000:4).

Naturalism is here characterised through use of the discipline criterion. It is, however, not a notion of the natural sciences that is supposed to do a lot of explanatory work in this definition. It is a notion of the empirical sciences in a certain state that is supposed to do this work.

At least three objections can be raised against this characterisation. Firstly, it needs to be specified what it means to say that an entity 'is somehow ontologically grounded in' certain objects. Secondly, invoking the notion of the empirical sciences raises the problem of specifying exactly when something is an empirical science. The worry is that the notion of the empirical sciences is just as vague as the notion of the natural sciences. And if the latter notion cannot be used to characterise naturalism because it is too vague, then the same must be true of the former notion. Thirdly, the notion of the hypothetically completed empirical sciences is problematic because it seems doubtful that the empirical sciences, even hypothetically, can be complete. What can be said in response to these objections?

As a specification of what they mean, Moser and Yandell say that it is either by reduction or supervenience that one entity can be ontologically grounded in another entity (ibid.:8). It should be noted here that if an entity only needs to stand in a supervenience relation with the objects countenanced by the hypothetically completed empirical sciences in order to be ontologically grounded in these objects (and therefore real), then naturalism becomes very inclusive. In philosophy of mind, the only position that is incompatible with the supervenience requirement, and which

therefore is not a version of naturalism, is Cartesian dualism that allows the mental to float free of the physical domain. A version of substance dualism that only differs from its Cartesian sibling, in that it insists that the mental cannot float free of the physical, is, however, a version of naturalism. In Moser and Yandell's characterisation, ontological naturalism is a world view that can accommodate the existence of entities outside space that are able to cause change in the physical world. 'Ontological naturalism' is of course just a name, and one can apply this name to whatever kind of world view one likes. If one, however, characterises ontological naturalism in such a way that it is expansive enough to allow for the existence of entities outside space that can causally interact with entities in space, then a very large number of theories qualify as a version of ontological naturalism, and not much is conveyed in terms of substantial information by saying that a certain theory is a version of ontological naturalism.

Regarding the second objection, it does not seem to me that the notion of the empirical sciences is just as vague as the notion of the natural sciences. What is distinctive of empirical science is that it formulates hypotheses that can be falsified or supported by observation of controlled experiments. Social sciences such as psychology, social psychology and anthropology formulate such hypotheses and therefore they belong to the class of empirical sciences. What about such a science as particle physics? One might argue that this is not an empirical science, but a highly theoretical science that puts forward abstract models for the nature and behaviour of various particles. If this is true, then the entities countenanced by this science (particles such as protons and neutrons) are not real. As a response to this, it can be argued that particle physics is an empirical science since many of its hypotheses can be tested against observation. Many of its hypotheses can be subjected to empirical

testing in the sense that they can be falsified or supported by observation of controlled experiments.

The third objection is quite powerful. The idea that the empirical sciences can be complete, rests, as far as I can see, on the following assumptions: i) the world contains a finite number of types of particles and a finite number of types of forces governing the behaviour of these types of particles. ii) the empirical sciences gradually become able to detect more and more of these types of particles and types of forces. iii) the empirical sciences will eventually detect all types of particles and all types of forces. When the empirical sciences have done this, they can be said to be complete. The history of the empirical sciences, however, does not lend much support to the idea that they will one day be complete.

The history of the empirical sciences is a one of continuous new discovery. What has very often happened is that what was taken to be the final truth about some matter has been revised in the light of further empirical investigation. The fact that the history of the empirical sciences is a history of continuous new discovery does not, of course, prove that the empirical sciences will continue to progress in this way and therefore this fact does not prove that the empirical sciences cannot be complete.

This is not, however, a problem for the person who is sceptical about the idea that the empirical sciences can be complete. I do not think that it is he who has to prove that the empirical sciences will continue to yield new information about the world. The defender of the notion of the hypothetically completed empirical sciences puts forward an idea that seems implausible in light of the history of the empirical sciences. I therefore think that it is reasonable to say that he must bear the burden of proving that the empirical sciences can be complete in the sense that one day they will have yielded all the information about the world that there is to yield. I find it

hard to see what a posteriori arguments can be used to support this claim, and I find it just as hard to see what a priori arguments can be used to support it.

There is another problem with the strategy of using the notion of the hypothetically completed empirical sciences in a characterisation of naturalism.

Consider this dialectic:

Q1: What is naturalism?

A1: Naturalism is the view that every real entity either consists of or is somehow ontologically grounded in the objects countenanced by the hypothetically completed empirical sciences.

Q2: What are the hypothetically completed empirical sciences?

A2: *The hypothetically completed empirical sciences* are the empirical sciences in such a state that they have countenanced every real object.

Q3: What is a real object?

A3: A real object is an object that is countenanced by *the hypothetically completed empirical sciences*.

We are trying to find out what naturalism is. We are then presented with a characterisation of naturalism that involves reference to the notion of the hypothetically completed empirical sciences. In order to get clear about what naturalism is, we therefore look for an elucidation of the notion of the hypothetically completed empirical sciences. We are then told that the hypothetically completed empirical sciences are the empirical sciences in such a state that they have countenanced every real object. When we then ask what a real object is, the answer we get involves reference to the notion of the hypothetically completed empirical sciences.

This is, however, a non-informative way of arguing. It was the notion of the hypothetically completed empirical sciences that we in the first place needed an elucidation of, in order to get clear about what naturalism is. The notion of the hypothetically completed empirical sciences is, however, now being explained in terms of itself. The definiendum is the notion of the hypothetically completed empirical sciences. The definiens must not contain the definiendum. However, this is what happens, somewhat disguised, in the above dialectic. To see this, consider the fact that the full answer to the question of what the hypothetically completed empirical sciences are, is the conjunction of A2 and A3. The full answer to Q2 is: 'the hypothetically completed empirical sciences are the empirical sciences in such a state that they have countenanced every real object, and a real object is an object that is countenanced by the hypothetically completed empirical sciences'.

The problem of arguing in a non-informative way is avoided if another answer than A3 is given. A3 is, however, the natural answer to give for one who has invoked the notion of the hypothetically completed empirical sciences in his characterisation of naturalism. If one does not give A3 as an answer, or does not give an answer that entails A3, then one allows for the possibility that an object should be real without it being the case that it is countenanced by the hypothetically completed empirical sciences. But if it is possible that an object should be real without it being the case that it is countenanced by the hypothetically completed empirical sciences, then there would not be much sense in saying that the empirical sciences are complete. For the sake of the argument, it can be agreed that the empirical sciences can be said to be complete without it being the case that they countenance everything real. There might be supervening entities, for example non-physical mental entities, that are real without it being the case that they are countenanced by the hypothetically

completed empirical sciences. If the empirical sciences are complete, they must, however, be able to countenance all the objects (atoms, molecules or other physical structures) that these supervening entities supervene upon. So, on pain of watering down the notion of the hypothetically completed empirical sciences, one is forced to give A3 as a reply to Q3. But then the problem of arguing in a non-informative way arises.

5. The Causal Criterion

Another way of explaining what naturalism is, as an ontological view, consists in making use of what can be called the 'causal criterion for a property's being natural' (Crisp 1996:115). If this criterion is adopted, then naturalism is the view that only natural properties are real, and that natural properties are those properties that pass the causal explanatory test. What is it for a property to pass this test? If a property is to pass this test, then it must be the case that it plays a role in the best causal explanation of phenomena that are in need of explanation. Our causal explanations of why events occurred contain reference to various entities and properties, and a given property passes the causal explanatory test if and only if the best causal explanation of at least one event, contains reference to this property. Properties that we never need to refer to in order to give the best causal explanation of events do not pass the causal explanatory test. Such properties are causally superfluous and therefore they cannot be said to be natural properties.

I do not think that the causal criterion for a property's being natural gives us any clear picture of what kind of general world view is entailed by naturalism. The causal criterion is silent on the important question of what the best causal explanation of an event is, and if it is silent on this question, then it does not make clear what

properties are natural. To characterise natural properties as the ones that are referred to in the best causal explanations is explanatory ineffective. Such a characterisation immediately raises the question of what the best causal explanations are. It is, moreover, controversial what properties have causal power. Consider, for example, the question of what the best causal explanation is of an agent forming a moral conviction in a given situation. On this question, one might side with Gilbert Harman and say that the best causal explanation of this phenomenon, or any other phenomenon for that matter, does not involve reference to distinct moral properties that supposedly are properties of the act, policy or institution with which the agent is confronted. The best causal explanation only involves reference to an external world that is fully describable in non-moral terms and to the agent's psychological make up.⁹

Alternatively, one might side with G.E. Moore and say that at least sometimes, the best causal explanation as to why an agent forms the moral conviction that a given act is intrinsically good, is an explanation that involves reference to an ontologically distinct property of goodness.¹⁰ For present purposes, no stand needs to be taken on the debate between Harman and Moore. In relation to a discussion of ontological naturalism and a discussion of various ways of characterising this position, it is, however, noteworthy that if naturalism is characterised through use of the causal criterion, then 'naturalism' or 'naturalist' are not predicates that cut much ice in terms of distinguishing one class of philosophical positions from a different class of philosophical positions. Any philosophical position will count as a version of naturalism, as long as it is an integral part of the

⁹ (Harman 1977:7).

¹⁰ (Moore 1993:111; 130).

position that the kind of entities that it postulates the existence of, feature in the best causal explanation of a phenomenon that demand causal explanation.

This is, of course, not to suggest that all causal explanations are equally well-founded. It is a central part of philosophical inquiry to assess the plausibility of the types of causal explanation that a given philosophical position offers of various phenomena, and a given philosophical position might be rejected on the ground that it offers an implausible type of causal explanation of a given phenomenon. The important project of assessing the plausibility of a type of causal explanation offered by a given philosophical position is, however, something, that comes after, and is independent of, the question of whether or not the philosophical position counts as a version of naturalism.

6. Methodological Naturalism

Where does all of this leave us? What I have tried to show in the previous sections, is that a number of popular characterisations of naturalism, as an ontological view, are uninformative, internally inconsistent or vague. This is not to say that a characterisation of ontological naturalism cannot be found that avoids these flaws. I shall not spend time trying to come up with a satisfying characterisation of naturalism as an ontological view. My reason for this is that I am not interested in defending naturalism as an ontological view, whatever exactly naturalism is as an ontological view. I do, however, think that methodological naturalism should be endorsed. What is methodological naturalism? One characterisation of this component of naturalism has already been presented. Here is another, and slightly more detailed, characterisation:

Core methodological naturalism: "every legitimate method of acquiring knowledge consists of or is grounded in the hypothetically completed methods of the empirical sciences (that is, in natural methods)" (Moser & Yandell 2000:10).

One problem here is the notion of the 'hypothetically completed methods of the empirical sciences'. The problem is of a structural kind similar to the one outlined in the discussion of the notion of 'the hypothetically completed empirical science'. To see that the structural problem is the same, consider this dialectic which has obvious similarities with the dialectic already considered:

Q1: What methods of acquiring knowledge belong to the set of methods that constitutes the hypothetically completed methods of the empirical sciences?

A1: A method of acquiring knowledge belongs to the set of methods that constitutes *the hypothetically completed methods of the empirical sciences*, if and only if the method of acquiring knowledge belongs to the set of methods that is employed by the empirical sciences when they are developed to such a level that they employ every legitimate method of acquiring knowledge.

Q2: What is a legitimate method of acquiring knowledge?

A2: A method of acquiring knowledge is a legitimate method of acquiring knowledge if and only if it is a method that belongs to the set of methods that constitutes *the hypothetically completed methods of the empirical sciences*.

The characterisation of methodological naturalism involves reference to the notion of the hypothetically completed methods of the empirical sciences, but this notion is explained, at least partly, in terms of itself. The full answer to the question of what methods of acquiring knowledge belong to the set of methods that constitutes the hypothetically completed methods of the empirical sciences, is this: 'a method of

acquiring knowledge belongs to the set of methods that constitutes the hypothetically completed methods of the empirical sciences, if and only if the method of acquiring knowledge belongs to the set of methods that is employed by the empirical sciences when they are developed to such a level that they employ every legitimate method of acquiring knowledge, *and* a method of acquiring knowledge is a legitimate method of acquiring knowledge if and only if it is a method that belongs to the set of methods that constitutes the hypothetically completed methods of the empirical sciences'.

The structural problem is avoided by giving a different answer to Q2 than A2. But if one does not give A2 as an answer here, or an answer that entails A2, then one allows for the possibility that there should be a legitimate method of acquiring knowledge that do not belong to the set of methods that constitutes the hypothetically completed methods of the empirical sciences. But if the set of methods that constitutes the hypothetically completed methods of the empirical sciences lack, or are without, a legitimate method of acquiring knowledge, then this set of methods cannot correctly be said to constitute the hypothetically completed methods of the empirical sciences. A set of methods can only correctly be said to be the set of methods that constitutes the hypothetically completed methods of the empirical sciences if it contains every legitimate method of acquiring knowledge. So, on pain of watering down the notion of the hypothetically completed methods of the empirical sciences, one is forced to give A2 as an answer to Q2. But then the problem arises that the notion of the hypothetically completed methods of the empirical sciences is, at least partly, explained in terms of itself. I take this argument to show that a notion of 'the hypothetically completed methods of the empirical sciences' should not occur in a characterisation of methodological naturalism.

Another problem with the above characterisation of methodological naturalism is the notion of a method being 'grounded in' a set of other methods. Moser and Yandell do not make clear what they mean by 'grounded in' and lacking any such clarification, the presence of the notion in a characterisation of methodological naturalism, makes it the case that methodological naturalism is a view that can encompass just about every conceivable method of inquiry. To take a familiar example, consider the laying of tarot cards as a method of acquiring knowledge. In some sense of 'grounded in', this is a method of inquiry that is grounded in the methods of the empirical sciences. The laying of tarot cards is a method of inquiry that presupposes theoretical assumptions. The deck of cards can be seen as experimental equipment and the process of laying the cards can be seen as the process of conducting an experiment. Furthermore, just as use of our perceptual faculties plays a vital role in the methodology of empirical science, use of our perceptual faculties plays a vital role in the method of acquiring knowledge that is constituted by the laying of tarot cards. One has to look at the cards in order to be able to form a view about what has happened, or what will happen.

If methodological naturalism is characterised in such a way that it encompasses a method of inquiry such as the laying of tarot cards, then methodological naturalism becomes a view that is too accommodating to be of much theoretical interest. The element in the characterisation of methodological naturalism that is responsible for methodological naturalism being so accommodating should therefore be removed. In short, a characterisation of methodological naturalism should not involve an unqualified notion of one method of inquiry being 'grounded in' the methods of other methods of inquiry.

These considerations suggest that Moser and Yandell's characterisation of methodological naturalism should be amended such as to yield the following characterisation of methodological naturalism:

Methodological naturalism: a method of acquiring knowledge is a legitimate method of acquiring knowledge if and only if it is a method employed by the best contemporary empirical science.

This is what I take methodological naturalism to be, and this is the methodological view that I think should be endorsed. My characterisation of methodological naturalism is not without problems. It is vague what empirical science is. I have, however, said something about what characterises empirical science and what sciences should count as empirical. It is also vague what the best contemporary empirical science is. There are many issues in the empirical sciences on which there is no agreement, and conflicting theories and explanations compete for allegiance. This does not, however, mean that the notion of the best contemporary empirical science is completely empty. It should be accepted that one empirical scientific theory is better than a competing theory if it accords better with the observational data, and if it enables a more coherent and systematic explanation of its subject matter than the competing theory. Newtonian physics does not, for example, constitute the best contemporary empirical science since there are other theories that yield a more coherent and systematic account of the subject matter of Newtonian physics. The fact that my characterisation of methodological naturalism contains vague notions means that it is itself vague. It yields, however, an approximate position whose plausibility can be the subject of debate.

Why should methodological naturalism be endorsed? I do not have any original arguments or insights to offer on this question. Methodological naturalism

should be endorsed because the empirical sciences have proved very successful in yielding coherent and systematic causal explanations of phenomena that we are interested in being able to causally explain. The empirical sciences have, moreover, proved able to deliver extremely accurate predictions of many types of future events in both micro- and macro-cosmos. It is, in short, the great explanatory and predictive success that the empirical sciences have enjoyed that warrants the endorsement of methodological naturalism. No other method of inquiry has been able to produce a catalogue of explanatory and predictive successes that is anywhere close to matching the catalogue of successes that the empirical sciences can put forward for inspection.

It should be stressed that methodological naturalism does not conceptually entail any particular view about what sort of entities constitute reality as a whole. The best contemporary empirical science makes reference to a wide variety of entities in its explanatory theories. The existence of some of these entities can be verified through use of our perceptual faculties. We have strong reason to believe in the existence of yet other entities because the assumption that these entities exist, enables a coherent and systematic explanation of a number of different phenomena that we are interested in being able to explain. The claim that reality contains nothing but these entities is, however, a claim that is distinct from the claim made by methodological naturalism, and it is not a claim that is conceptually entailed by methodological naturalism.

Chapter 3

Supervenience

1. *Introductory*

The general topic of this chapter is the concept of supervenience. In sections two to four, I outline and discuss a number of common types of supervenience. In section five, I argue that a particular definition of weak supervenience, put forward by Kim, should be refined in a certain way. I also argue that a particular definition of weak supervenience, put forward by Blackburn, should be rejected, and that the relation of supervenience should be taken to be no more than a relation of necessary co-variation. In section six, I try to make clear why supervenience, understood merely as a relation of necessary co-variation, is still an interesting concept in ethics. I also suggest a particular definition of moral supervenience, and argue that anyone who accepts moral supervenience in this definition, incurs an explanatory debt that must be discharged.

2. *An Initial Characterisation*

The verb 'supervene' is part of everyday language. It derives from the Latin 'super' which means 'on', 'above' or 'additional', and from the Latin verb 'venire' which means 'to come'. The *Oxford Advanced Learner's Dictionary* defines 'to supervene' as 'to occur as an interruption or change' as in 'she worked well until illness supervened'.¹¹ The philosophical use of 'supervene' is, however, different from the way the term is used in everyday language. 'Supervene' and its derivatives

¹¹ Fourth edition, Fourth Impression. Oxford: Oxford University Press 1990.

'supervenient' and 'supervenience' are, when used in philosophical literature, technical terms that stand in no obvious connection to the uses they have outside philosophy. This means that if in philosophical literature something is said to supervene upon something else, this is not supposed to mean that it is something that 'occurs as an interruption or change'.

One way of giving an initial characterisation of the philosophical concept of supervenience is to quote David Lewis:

"The idea is simple and easy: we have supervenience when [and only when] there could be no difference of one sort without differences of another sort".¹²

The philosophical concept of supervenience is standardly taken to denote a dependence relation of a modal sort between distinct sets of entities: the supervening entities (A-entities) and the subvenient or base entities (B-entities). If A-respects supervene upon B-respects, then the former is dependent on the latter in the sense that there cannot be a variation in A-respects without a variation in B-respects. Variation in the supervening A-respects requires, that is, variation in the subvenient B-respects. A useful metaphorical picture to invoke in order to characterise the dependence relation between A- and B-respects in cases where the former supervene upon the latter, consists in saying that A-respects cannot 'float free' of B-respects.

To fix the concept of supervenience further in mind, an example might be helpful. It is a widely held, but not uncontroversial, view that mental respects supervene upon physical respects. According to the simple idea of supervenience under consideration, this view is true if and only if there cannot be a difference in mental respects without a difference in physical respects. If such a supervenience relation holds between mental and physical respects, then it is true that identity in

¹²(Lewis 1986:14)

terms of physical respects entails identity in terms of mental respects. If there cannot be a difference in mental respects without a difference in physical respects, then any difference in the former respect requires a difference in the latter respect. A slightly different characterisation of the simple concept of supervenience is therefore this: A-respects supervene upon B-respects iff it is necessary that if there is no difference in B-respects, then there is no difference in A-respects.

Supervenience can be construed as both a synchronic and diachronic relation. A supervenience relation of the former kind is one that involves the A- and B-respects of at least two objects. If there cannot be a difference in the A-respects of two objects without there being a difference between them in terms of B-respects, then, and only then, is there a synchronic supervenience relation of A-respects on B-respects. A supervenience relation of the latter kind is one that involves the relation through time of the A- and B-respects of a single object. A diachronic supervenience relation implies that an object cannot change in terms of some A-respect from time t to time t^* without changing in terms of some B-respect from time t to time t^* . Defenders of psycho-physical supervenience are likely to endorse both these supervenience relations. Their claim is that two persons cannot differ in terms of mental state without differing in terms of physical constitution, and that a single person cannot change over time in terms of mental state without changing over time in terms of physical constitution.

If the core idea of supervenience is the idea of a dependence relation of a modal sort, then we can ask the question of what the relata can be of this relation. Between the members of what ontological categories can a supervenience relation obtain? Up until now the relata of the supervenience relation has been said to be either A- and B-entities or A- and B-respects. This is quite vague since a number of

different things can reasonably be said to fall under the concept of an entity, and to say that two things differ in terms A-respects is compatible with it being the case that they differ in terms of what A-properties they instantiate, what the A-facts are about them and what A-states of affairs obtain in them. The generality of the terms involved in this initial characterisation of supervenience is, however, not something that is to be deplored. It namely makes possible a characterisation of supervenience that is compatible with, or makes allowance for, the fact that a supervenience relation can obtain between the members of a number of different ontological categories.

It is perhaps most common to present supervenience as a modal dependence relation between ontologically distinct sets of properties.¹³ As an example of property supervenience consider the thesis that chemical properties supervene upon physical properties. If this thesis is true, then there cannot be a difference in terms of what chemical properties two substances instantiate without there being a difference in terms of what physical properties they instantiate. For example, if only one of two substances is an acid, then the two substances cannot be identical in terms of molecular constitution. Any difference at the level of chemical properties requires a difference at the level of physical properties.

However, facts of one sort can supervene upon facts of another sort, and predicates of one sort can supervene upon predicates of another sort. As an illustration of a supervenience thesis in terms of facts, consider the very plausible claim that facts about being a prime number supervenes upon facts about being divisible by certain numbers. Two numbers cannot differ in terms of facts about being a prime number without differing in terms of facts about being divisible only by themselves and by 1. If it is a fact about one of two numbers that it is a prime

¹³See (Kim 1984:55; 1990:140)

number, and a fact about the other number that it is not a prime number, then it cannot be a fact about each of the two numbers that it is only divisible by itself and by 1.

As an example of a supervenience thesis in terms of sets of predicates, consider the thesis that evaluative predicates supervene upon non-evaluative predicates. There cannot be a difference in terms of what evaluative predicates are true of two objects without there being a difference in terms of what non-evaluative predicates are true of them. Likewise, it is necessary, that if there is a change over time in terms of what evaluative predicates are true of a given object, then there is a change over time in terms of what non-evaluative predicates are true of the object.

The above supervenience theses are in accordance with Lewis' characterisation of the core idea of supervenience. Despite the fact that the three theses carve out in different ways, Lewis' talk of a difference of one sort and differences of another sort, they all conform to the simple and essential idea of supervenience that there cannot be a difference of one sort (a difference at the supervening level) without differences of another sort (differences at the subvening level).

3. Types of Supervenience

In the literature, it is common to distinguish between three different types of supervenience: namely weak (intra-world), strong (inter-world) and global supervenience.¹⁴ Each type of supervenience has been given a number of different definitions by different authors. For the purpose of simplicity, I shall here confine

¹⁴ See (Kim 1984; 1987; 1990), (Blackburn 1985a), (Jackson 1998) and (McLaughlin 1995).

myself to presenting and discussing Kim's (1987) definitions of the three types of supervenience.

If A and B denote two sets of ontologically distinct properties, and if A is the supervening set and B the subvening set, then weak supervenience can be given the following definition: A-properties weakly supervene upon B-properties iff:

(WS) "Necessarily, for any object x and any property F in A, if x has F, then there exists a property G in B such that x has G, and if any y has G, it has F" (Kim 1987:80).¹⁵

This passage specifies the necessary and sufficient condition that must obtain in order for A-properties to weakly supervene upon B-properties. (WS) is equivalent to the claim that it is impossible that there should exist two things, x and y, such that x and y are indiscernible in respect of properties in B and yet discernible in respect of properties in A. This claim is, in turn, equivalent to the claim that it is necessary that if two things, x and y, are indiscernible (or identical) in respect of properties in B, then they are indiscernible in respect of properties in A.

Weak supervenience of A-properties upon B-properties asserts a certain relation between A-properties and B-properties of objects within a given world. This is the reason for calling it intra-world supervenience. Consider, for example, the thesis that mental properties weakly supervene upon physical properties. If this thesis is true, then it is true in every possible world, W, that if two creatures are identical in terms of physical properties, then they are identical in terms of mental properties. This means that if some person in our world has physical properties G* and mental property F, then any person in our world who has physical properties G* has mental property F as well.

¹⁵ Put formally: A-properties weakly supervene on B-properties iff:
 $\Box \forall x \forall F \in A [F x \rightarrow \exists G \in B (G x \ \& \ \forall y (G y \rightarrow F y))]$.

The thesis that mental properties weakly supervene upon physical properties is, however, compatible with it being the case that G*-creatures in other possible worlds, including worlds completely identical to our world in terms of physical properties, do not have mental property F. In a world that is a complete physical duplicate of our world, W*, G*-creatures may have mental property F*, or they may indeed have no mental properties at all. Weak supervenience of A-properties on B-properties only requires that within W*, mental properties are distributed consistently across G*-creatures such that G*-creatures are identical in terms of whatever mental properties they instantiate. The significant feature of weak supervenience is simply that it allows that two objects in different worlds can be B-identical without being A-identical.

Keeping the above definition of A and B in mind, strong supervenience can be given the following definition: A-properties strongly supervene upon B-properties iff:

- (SS) "Necessarily, for any object x and any property F in A, if x has F, then there exists a property G in B such that x has G, and necessarily if any y has G, it has F" (ibid.:80).¹⁶

Strong supervenience of A-properties upon B-properties is stronger than weak supervenience in that it requires B-identical objects in different worlds to be A-identical. If mental properties strongly supervene upon physical properties and if there is a creature in some world, perhaps our own world, that has physical properties G* and mental property F, then it is true for any creature, not only in that particular world but in any possible world, that if this creature has physical properties G*, then it has mental property F. Strong supervenience of A-properties on B-properties

¹⁶ Put formally: A-properties strongly supervene on B-properties iff:
 $\Box \forall x \forall F \in A [Fx \rightarrow \exists G \in B (Gx \ \& \ \Box \forall y (Gy \rightarrow Fy))]$.

guarantees stability across worlds for the correlation between specific A-properties and their base properties.

Kim actually presents the above two definitions of weak and strong supervenience as definitions of weak and strong covariance. It is, however, not problematical to cite the two definitions as definitions of weak and strong supervenience since Kim acknowledges that using the concept of covariance and not the concept of supervenience is a purely terminological decision. It should, however, be noted that supervenience is not the same concept as covariance. The two concepts should therefore not, without explanation, be used interchangeably. There are a number of different ways in which A-properties can covary with B-properties, and if A-properties supervene upon B-properties, then they covary with them in a particular way. A-properties can covary with B-properties in such a way that it is necessary that if there is a variance in the B-properties of an object, then there is a variance in its A-properties. But this relation of covariance between A-properties and B-properties is significantly not the same as a supervenience relation of A-properties on B-properties.

An important question in connection with both weak and strong supervenience is the question of what the strength is of the modal term(s) involved in a supervenience thesis (or alternatively, what the range is of the world-binding quantifier(s) involved in a supervenience thesis). Standard options here include physical, metaphysical and conceptual necessity.¹⁷ In general definitions of the theses of weak and strong supervenience, as the ones above, no answer has to be given to the question of what the exact strength is of the modal term(s). Supervenience theses can, depending upon what sets of properties are denoted by 'A' and 'B', hold with

¹⁷See (Kim 1990:141).

various kinds of necessity, and it is useful to have general definitions of supervenience that allow for this fact.

Let us now turn to global supervenience. In global supervenience, the general principle that there could be no difference of one sort without differences of another sort is taken to hold for whole worlds and not for objects within either one or more possible worlds. What global supervenience of A-properties on B-properties roughly asserts is that two worlds cannot differ in terms of A-properties without differing in terms of B-properties.¹⁸ A more precise definition of global supervenience is this: A-properties globally supervene upon B-properties iff:

(GS) "Any two worlds indiscernible with respect to B-properties are indiscernible with respect to A-properties" (Kim 1987:82).

If this definition of global supervenience is to be of any interest, then we need to know what it is for two worlds to be indiscernible with respect to a particular set of properties. McLaughlin notes that by worlds that are indiscernible (i.e., twins) with respect to A- [or B-] properties is typically meant worlds that have the same total pattern of distribution of A- [or B-] properties.¹⁹ This is informative, but now we want to know what it exactly means for two worlds to have the same total pattern of distribution of A- [or B-] properties. A suggestion by Kim is that two worlds, *W* and *W**, are indiscernible with respect to B-properties if and only if for any B-property *G* and any individual *x*, *x* has *G* in *W* if and only if *x* has *G* in *W** (Kim 1984:68). If some individual in *W* has some B-property *G*, then this individual's counterpart in *W** also has *G* and vice versa. This means that there is no individual in either *W* or *W** that has a B-property that its counterpart in the other world does not have. The

¹⁸ See, for example, (Kim 1984:68) and (McLaughlin 1995:31).

¹⁹ (McLaughlin 1995:31).

number of B-properties in the two worlds is identical, and the B-properties are distributed across the same number of individuals in the same way in the two worlds.

It has been argued that on Kim's proposal, two worlds, W and W^* , can be B-indiscernible even if they do not have the same number of individuals.²⁰ B-indiscernibility between worlds that do not have the same number of individuals just requires that the individual(s) that exists in only one of the two worlds does not instantiate any B-properties. Consider two worlds, W and W^* , that are such that the former has two individuals, x and y , and the latter three individuals, x , y and z . Assume, moreover, that whatever B-property, or set of B-properties, x and y respectively have in W is a B-property, or set of B-properties, that x and y respectively have in W^* , and that z in W^* does not instantiate any B-property.²¹ This may constitute an example of a pair of worlds that are B-indiscernible despite the fact that they do not have the same number of individuals. The reason for thinking that W and W^* are B-indiscernible in an example such as this is that it is true of no individual in either W or W^* that it instantiates a B-property that is not instantiated by its counterpart in the other world. W and W^* contain the exact same number of B-properties and these B-properties are distributed in an identical way across an identical number of individuals in the two worlds.

It seems to me that Kim has presented us with an intuitive compelling account of what B-indiscernibility between two worlds consists in, and that this account is compatible with it being the case that two worlds can be B-identical in two distinct ways: one in which there is an identical number of individuals in the worlds under consideration, and one in which there is a non-identical number of individuals in the worlds under consideration. One might, however, still be sceptical about the

²⁰ (McLaughlin 1995:32).

²¹ Let us call this pair of worlds 'LL'.

idea that two worlds can be B-identical if there is not an identical number of individuals in the two worlds. Let me first try to give an argument for this sceptical view and then try to show that this argument fails.

The argument in favour of the sceptical view can be spelled out in relation to the pair of worlds, LL, discussed above. The structure of the argument is that z in W^* actually does instantiate a B-property and since z has no counterpart in W there is no individual in W that instantiates the B-property that z instantiates in W^* . There is therefore a B-property instantiated in W^* that is not instantiated in W , and the two worlds are therefore B-discernible. What B-property is it that z supposedly instantiates? If the set of B-properties is a non-empty set, then there is a B-property that is the conjunction of all the specific B-properties: namely $b_1 \ \& \ b_2 \ \& \ b_3 \ \& \dots \ b_n$ and if there is such a property, then there is also a conjunctive B-property of the form $\neg b_1 \ \& \ \neg b_2 \ \& \ \neg b_3 \ \& \dots \ \neg b_n$, and it is this B-property that z instantiates.

I think that a reductio can be used to show that this argument is not very convincing. Consider a world that is characterised by the fact that it contains physical substances and therefore physical properties (B-properties) and non-physical substances and therefore non-physical properties (A-properties). We can just assume that the world in question contains angels, and that the property of being an angel is a non-physical property. Now, the above argument allows the conclusion that the angels in that world have a physical property: namely the property of not being $b_1 \ \& \ b_2 \ \& \ b_3 \ \& \dots \ b_n$ where this set of properties is the set of instantiated physical properties in that world. A paradigm example of something non-physical, an angel, would actually have a physical property. This consequence of the argument seems to me to be rather absurd, and I therefore think that the argument should be rejected. The same conclusion should be reached if we consider another application of the

argument. It seems that the argument can be used to show that particulars in our world such as cups, tables and indeed subatomic particles have a mental property: namely the property of not being a_1 & a_2 & a_3 &..... a_n where this is the set of mental properties actually instantiated in our world. But the idea that, say, cups and electrons should have a mental property is, in my mind, not one that should be taken seriously.

It is a significant feature of global supervenience that it does not entail weak supervenience (Kim 1987:83). The former type of supervenience can, that is, hold when the latter type does not hold. To see this, consider a simple world, W , in which only two individuals, x and y , exist. Imagine, moreover that x has B-property G and A-property F and that y has G but not F . Such a world is excluded by the truth of the claim that A-properties weakly supervene upon B-properties. The existence of such a world is, however, compatible with it being the case that A-properties globally supervene upon B-properties. The only thing that is required by the global supervenience of A-properties on B-properties is that there is not another world, W^* , that is B-identical to W (i.e. a world in which G is distributed over the same individuals as in W) but non-identical to W in terms of how specific A-properties are distributed across individuals. The existence of a world in which only individuals x^* and y^* exist and in which x^* is G and F , and y^* is G and F is the kind of world that, provided the existence of W , would prove wrong the claim that A-properties globally supervene upon B-properties.

Kim has argued, quite convincingly in my mind, that the feature of global supervenience that it allows what weak supervenience excludes is a feature that makes it difficult to see how global supervenience can qualify as a significant dependency relation between two sets of properties (Kim 1990:155). The line of thought is that it can hardly be the case that one set of properties (the A-properties)

are dependent upon or determined by another set of properties (the B-properties) if it is permitted, as it is by global supervenience of A-properties on B-properties, that two objects within a given world can be B-identical without being A-identical. To see this more clearly, consider a thesis of global psycho-physical supervenience in which mental properties are taken to be A-properties and physical properties are taken to be B-properties. Such a thesis permits that there could exist a person in this world that is a complete physical replica of me but differs from me in terms of having no mental life at all. The fact that global psychophysical supervenience permits the existence of such a person should lead us to follow Kim, and conclude that this supervenience thesis is incompatible with any idea to the effect that the mental or psychological nature of the world is somehow dependent upon or determined by the physical nature of the world.

The view that global supervenience is too weak to sustain a dependency relation of significance between two sets of properties receives support from yet another consideration. Global supervenience of A-properties on B-properties permits that worlds can differ in the smallest of details in their B-properties and yet differ very substantially in terms of their A-properties. Global supervenience of mental properties on physical properties permits, for example, that there can be a world, W^* , that differs from our world in that physical respect that the Pacific ocean in W^* contains one more H_2O molecule than our Pacific ocean, and differs from our world in that mental respect that every person in W^* is constantly joyful or perhaps in that mental respect that in W^* it is only rocks of a certain kind that instantiate consciousness. Global supervenience of the mental upon the physical in effect says that if there is some difference in physical respects between two worlds, then there is no limit to the extent to which the two worlds can differ in terms of mental respects.

4. *Ascriptive Supervenience*

A useful distinction has been drawn between 'ontological' and 'ascriptive' supervenience.²² Ontological supervenience is a relation between classes of properties and theses of weak, strong and global supervenience are all compatible with ontological supervenience. The only thing that this requires is that the relata of these different types of supervenience theses are classes of properties. Endorsing ontological supervenience in a certain area of discourse commits one to realism about the supervening class of properties. It is, for example, only intelligible to hold the view that moral properties supervene upon naturalistic properties if one recognises the existence of moral properties. Otherwise, one would hold the paradoxical view that a certain class of properties, that one does not recognise the existence of, supervenes upon another class of properties.

Ascriptive supervenience is, on the other hand, a relation between classes of judgements (e.g. moral judgements and judgements ascribing naturalistic properties to objects) and endorsing this kind of supervenience in a certain area of discourse, does not commit one to realism about the supervening class of properties. In Klagge's words:

"Ascriptive supervenience is the view, roughly, that, logically speaking, a person's judgements of a certain (supervening) kind about things cannot differ unless judgements of the other kind about the things differ" (Klagge 1988:462).

Applied to ethical discourse, ascriptive supervenience is the view that a person's moral judgements about two objects cannot differ unless his judgements about the naturalistic properties of the two objects (his naturalistic judgements about the two

²² See (Klagge 1988).

objects) differ. The underlying intuition here is that it is part of the meaning of moral terms that judgements involving them must supervene upon naturalistic judgements. If one puts forward identical naturalistic judgements about objects x and y, and then puts forward non-identical moral judgements about them, then one has misunderstood a part of the meaning of the moral terms that are involved in one's moral judgements. This means that the 'cannot' occurring in the above definition of ascriptive supervenience denotes conceptual impossibility.

R. M. Hare is a classic advocate for ascriptive supervenience. Consider this passage:

"First, let us take that characteristic of 'good' which has been called its supervenience. Suppose that we say 'St. Francis was a good man'. It is logically impossible to say this and to maintain at the same time that there might have been another man placed in exactly the same circumstances as St. Francis, and who behaved in exactly the same way, but who was different from St. Francis in this respect only, that he was not a good man" (Hare 1952:152).

Supervenience is here understood to be a characteristic of the term 'good'. Hare's view is that anybody who judges St. Francis to be a good man is forced by a part of the meaning of the term 'good' to give the same judgement of any other man judged to be naturalistically identical to St. Francis. Supervenience is, according to Hare, a conceptual constraint on any user of moral terms. His remarks about supervenience do not imply that it is logically impossible that there should be a naturalistic duplicate of St. Francis that differs from St. Francis in that he is not a good man. For this to be the case, there only needs to be a different user of moral terms that has different moral principles, and is of the opinion that the naturalistic properties

instantiated by St. Francis warrants the judgement that he is not a good man. This other user of moral terms is of course also bound by the conceptual constraint of supervenience. He is therefore forced, on pain of committing a conceptual mistake, to give an identical moral judgement of any other person he judges to be yet another naturalistic duplicate of St. Francis.

5. A Formal Feature of Weak Supervenience

After this introduction to the general concept of supervenience, and a survey of the most common types of supervenience theses, I now turn to consider a formal feature of weak supervenience. Consider the suggestion that, say, aesthetic properties supervene on natural properties in the way defined by (WS). Consider also a philosophical dualist who thinks that aesthetic properties stand in no systematic relation to natural properties. In particular, he thinks that aesthetic properties can float free of natural properties. Can such a dualist agree that aesthetic properties weakly supervene on natural properties? Due to a particularity of (WS), it actually seems that he can. For any aesthetic property F, (WS) is (trivially) true if G is taken to be a natural property that it is impossible that two numerically distinct objects should both instantiate. If it is impossible that two numerically distinct objects should both be G, then it is (trivially) true that if one object is F and G, then any object that is G is F. So, (WS) is trivially true in all cases where something is F and G, and it is impossible that two numerically distinct objects should both be G. A dualist of the above kind can exploit this point, and say that whenever an object instantiates some F, then the natural property G it also instantiates, is such a complex property that no

other object could instantiate it. He is therefore entitled to say that aesthetic properties weakly supervene on natural properties.²³

One might be of the opinion that this formal feature of the weak supervenience thesis makes it the case that the original definition of weak supervenience is in need of refinement. If anybody can agree that aesthetic properties weakly supervene on natural properties, then a claim to the effect that this relation holds, seems to become insignificant and of no interest. We want, however, such a claim to be of interest. It is supposed to say something interesting about the relation of two sets of properties. I share this opinion, so I think Kim's original definition should be refined by stipulating that G must be a property that it is possible that more than one object should be able to instantiate. By stipulating this, one avoids the above problem.

One might be dissatisfied with Kim's definition of weak supervenience for a different reason. This dissatisfaction is based on the idea that there is a consequential aspect embedded in the philosophical concept of supervenience. The idea is that when one property of an object supervenes upon another property of the object, then the object instantiates the supervening property in virtue of instantiating the subvening property. The supervening property F is in a sense consequential upon the subvening property G. The object has F in virtue of having G. Consider an instance of moral supervenience. If it is said that St. Patrick's property of being a good man supervenes upon a set of his naturalistic properties, then, on this suggestion, at least part of what is meant by this is that St. Patrick is a good man in virtue of, or because

²³ To take just one example: the dualist about aesthetics could say that what the elegance (F) of a particular building weakly supervenes upon is a natural property (G) that includes the property that the building instantiates in virtue of occupying a certain spatio-temporal position. It is both physically and metaphysically impossible that two numerically distinct buildings (two physical structures) should occupy the same spatio-temporal position. It is therefore trivially true that F weakly supervenes on G.

of, this set of naturalistic properties. His instantiation of the property of being a good man is, so to speak, a consequence of his instantiation of a set of specific naturalistic properties.

Let me try to spell out why this suggestion as to what is involved in our ordinary philosophical concept of supervenience is important for the question of whether or not the original definition of weak supervenience should be refined. If there is a consequential aspect embedded in our ordinary philosophical concept of supervenience, then, importantly, the relation of supervenience is not just a relation of necessary co-variation. To say that something supervenes upon something else is not just to say that there cannot be a variation at the supervening level without a variation at the subvening level. The thesis of necessary co-variation is an essential part of the concept of supervenience, but there is a further consequential aspect to the concept of supervenience. If there is more to the supervenience relation than it just being a relation of necessary co-variation, then it is important to note that (WS) cannot be used in a definition of supervenience if we by 'definition' mean a specification of the necessary and sufficient conditions that must obtain in order for a concept to apply. The reason for this is that (WS) only specifies a relation of necessary co-variation. It does not capture, or say anything about, the consequential aspect of the supervenience relation. This is a point very neatly made by John Heil:

"Indeed, (WS) implies only a kind of property correlation or co-variation. [...]. When a 's supervene on β 's, a 's are in some way dependent on or determined by β 's. When this is so, it is tempting to say that an object that possesses an a does so in virtue of possessing some β . This, however, would be to go beyond (WS). [...]. (WS) is a purely modal notion" (Heil:148).

(WS) only specifies the necessary condition that must obtain in order for A-properties to weakly supervene upon B-properties. It does not specify the sufficient condition because there is more to the supervenience relation than it simply being a relation of necessary co-variation. In order to capture the full meaning of the supervenience relation, i.e. the 'in virtue of' requirement that is embedded in the relation, something needs to be added to (WSF). The original definition of supervenience that involves reference to (WSF) is therefore in need of refinement.

It seems to me that if the assumption of this argument, the claim that there is a consequential aspect involved in our ordinary philosophical concept of supervenience, is accepted, then the argument is quite strong. It shows rather conclusively that the original definition of weak supervenience is in need of refinement. Let us, for the moment, accept the assumption. We can now ask the question of what a formula should look like if it is to give the necessary and sufficient condition that must obtain in order for one property to supervene upon another. In his 1985a work, Blackburn offers this suggestion:

$$(S) \quad N [(\exists x)(Fx \ \& \ G^*x \ \& \ (G^*x \ U \ Fx)) \rightarrow (\forall y)(G^*y \rightarrow Fy)]$$

This formula says: necessarily, if there is something that is F and G* and whose G*-ness underlies its F-ness, then anything G* is F. Blackburn's refined formula is an attempt to overcome the objection that (WS) only specifies a relation of necessary co-variation. In his 1985a work, Blackburn is explicitly of the opinion that there is a consequential aspect embedded in the ordinary concept of supervenience and he thinks that the refined formula captures this aspect.

"Belief in supervenience is then at least the belief that whenever a thing is in some F state, this is because it is in some underlying G state, or it is by virtue

of its being in some underlying G state. This is the minimal sense of the doctrine" (Blackburn 1985a:131).

(S) involves a rather uncommon sentence-operator: namely, U. This operator is in need of clarification. What does it exactly mean that one set of properties underlies another property? One cannot make use of (S) in a definition of this relation. One cannot, that is, define the underlying relation by saying that G^* underlies F if and only if (S). But how should we then understand the underlying-relation? Blackburn does not explain to us what the truth conditions are of an expression involving the U-operator. We know that the expression ' $G^*x \ \& \ Fx$ ' is true just in case both of the conjuncts are true, and because we know the truth conditions of this expression, it can usefully be used in a formula that is part of a definition of a particular concept. Because it is unclear what the truth conditions are of the expression ' $G^*x \ U \ Fx$ ', it is difficult to see how a formula involving this expression can cast light on the concept of supervenience.

These considerations about the U-operator should in my opinion make us believe that Blackburn does not, by invoking the formula (S), succeed in specifying the necessary and sufficient condition that must obtain in order for one property to supervene upon another. (S) simply cannot do the work that it is supposed to do.

Of course, the discussion of a single case does not allow one to draw a general conclusion, but it is my contention that if the supervenience relation is taken to contain a consequential aspect, the 'in virtue of' requirement, then the supervenience relation cannot be given a clear definition involving standard logical notation. Any definition of the form: 'A-properties supervene upon B-properties iff (S)', where (S) is a formula involving standard logical notation will fail to capture the consequential aspect of supervenience. If I am right about this, is it then the case that

the concept of supervenience is not an interesting philosophical concept? Not necessarily. There is the possibility, one that I think should be endorsed, of just taking supervenience to be a relation of necessary co-variation. If supervenience is taken to be such a relation, then we have a well-defined concept. (WS) can be used to give the necessary and sufficient condition that must obtain in order for one class of properties to (weakly) supervene upon another class. Is such a narrow supervenience concept of any philosophical interest?²⁴ It depends crucially on whether or not a relation of necessary co-variation of the kind expressed by (WS) is ever instantiated. If it is ever instantiated, then there is the task of explaining why it is the case that there cannot be a variation at the supervening level without variation at the subvening level.

6. *Supervenience and Ethics*

Let me end this chapter by saying something about why I think that the concept of supervenience is an interesting concept in the realm of ethics. It strikes me as very plausible that some sort of supervenience relation holds, in the narrow sense of supervenience, between moral properties and natural properties. Such a commitment to supervenience in the realm of ethics is something that is very widely found in the contemporary metaethical literature.²⁵ If there is no relation of supervenience, in any

²⁴It would be to confuse matters to use the term 'weak' of a concept of supervenience that only involves a relation of necessary co-variation and the term 'strong' of a concept of supervenience that in addition to this element involves a consequential aspect. As previously shown, these terms are already in use to mark a distinction of a quite different kind between concepts of supervenience. I therefore suggest that the term 'narrow' should be used of a supervenience concept that only involves the relation of necessary co-variation and that the term 'broad' should be used of a supervenience concept that has the additional aspect.

²⁵ "Everyone agrees that the moral features of things supervene on their natural features [...]. That is, everyone agrees that two possible worlds that are alike in all of their natural features must also be alike in their moral features" (Smith 1994:21), "By all accounts it is a conceptual truth that the moral features of acts supervene on their naturalistic features:" (Smith 2000:24), "The most salient and least controversial part of folk moral theory is that moral properties supervene on descriptive properties, that the ethical way things are supervenes on the descriptive way things are" (Jackson 1998:118), "In the case of ethics, supervenience seems to be built into the discourse - it is analytic - and I still see

of the different senses of supervenience discussed in this chapter, between moral properties and natural properties, then it is possible that moral properties should float free of natural properties. It is, in other words, possible that two objects, two states of affairs or two (entire) worlds should differ in terms of moral properties without differing in terms of natural properties. I do not consider this to be a genuine possibility. Assuming that there is some sort of supervenience relation between moral properties and natural properties, there is the task of identifying and describing this relation in detail. If and when this has been done, there is, however, also the further task of explaining why this relation holds. There is, in other words, the task of explaining what grounds the supervenience relation.²⁶

How should moral supervenience be characterised? It seems to me that at least weak supervenience is appropriate here. Minimally, what one is committed to just in virtue of participating in everyday moral discourse and practice is that there is no possible world in which two (or more) things differ in terms of moral properties without differing in terms of natural properties (or, if one is a non-descriptivist, there is no possible world in which an agent's moral judgements about two (or more) objects differ without it being the case that his non-moral judgements about them differ). The kind of necessity involved in moral supervenience should be construed as conceptual necessity. One commits a conceptual mistake if one deems it possible that two objects within the given world should differ in terms of moral properties without differing in terms of natural properties. Since global supervenience does not entail weak supervenience, one should not characterise moral supervenience merely in

quasi-realism as the only genuine explanation of this " (Blackburn 1998:315), "Moral assessment, virtually everyone concedes, supervenes upon the natural world in the following sense: two situations can differ in their moral qualities just in case they also differ in some natural qualities" (Railton 1995:99).

²⁶ The notion of 'grounding supervenience' is borrowed from (Heil:150).

terms of a thesis of global supervenience. In my view, it is correct to say that moral properties globally supervene on natural properties, but any such claim is potentially misleading, if it is not conjoined with a claim to the effect that moral properties also weakly supervene on natural properties. In the following, I shall take moral supervenience to be a relation of weak supervenience. More precisely, I shall construe moral supervenience as the claim that moral properties (A-properties) are related to B-properties (natural properties) such that:

Necessarily, for any object x and any property F in A , if x has F , then there exists a property G in B such that x has G , and if any y has G , it has F .²⁷

It is one thing to make a claim to the effect that a supervenience relation exists between moral properties and natural properties, but if this claim is not backed up by an explanation as to why this relation holds, then the claim amounts to a brute postulate. Anyone who accepts moral supervenience in my definition, and who wants to be on philosophical safe ground has to be able to explain why this relation holds. Acceptance of moral supervenience does not come free of charge. By accepting it, one always incurs an explanatory burden.

Now, it might be the case that not all metaethical theories are equally good at explaining why moral supervenience holds. Moreover, if it is agreed that moral supervenience does hold, and that this relation is something that is in need of explanation, then it is not a harmless feature of a particular metaethical theory that it is not able to explain why moral supervenience holds. It seems to me that any minimally convincing metaethical theory must be able to explain why moral supervenience holds. If it is the case that a particular theory has no explanation of supervenience, then this feature of it constitutes a reason for simply dismissing the

²⁷ The modal term here expresses conceptual necessity, and it is implicitly assumed that G is a property that more than one object can instantiate.

theory, and if one metaethical theory has a better explanation of moral supervenience than its competitors, then this should be seen as something that gives us a reason for preferring this theory over its competitors. It is not a conclusive reason because even though it is important to fare well on the issue of moral supervenience, faring well on this particular issue is not the only thing that is important to the overall plausibility of a metaethical theory.

Chapter 4

Realism, Expressivism and Moral Supervenience

1. *Introductory*

In this chapter, I try to bring out in more detail how considerations about supervenience can be used to evaluate metaethical theories. I shall be concerned with two such theories, and I will look at how well each of these is able to explain moral supervenience. In sections two and three, I outline, refine and discuss Simon Blackburn's influential supervenience argument against a particular type of moral realism. I argue that this argument is powerful, and that the type of moral realism that Blackburn considers is problematic because it cannot offer a convincing explanation of moral supervenience. In section four, I focus on the semantic tenet of Blackburn's quasi-realist position. In addition to giving an exposition of the key features of an expressivist analysis of moral utterances, I try to make clear what might motivate such an analysis. In section five, I discuss the expressivist explanation of moral supervenience. The reason for doing this is that the issue of moral supervenience only provides reason to prefer expressivism over the realist position considered by Blackburn, if the former can give a convincing explanation of moral supervenience. I argue that expressivism can offer such an explanation, and that it does not have an independent problem on the issue of moral supervenience.

2. Blackburn's Argument

Consider a position in metaethics that might be labelled 'Moorean realism'. This position essentially involves the claims that there are instantiated moral properties, and that these properties belong to their own ontological class. Moorean realism therefore advances pluralism about properties. There is a class of natural properties and there is a class of moral properties, and properties from the latter class cannot be reduced to properties from the former class (or to properties from any other class).

Moorean realism can be further specified by saying that it involves the claim that there is no conceptual connection between an object having certain natural properties and its having certain moral properties. Let x be an object, let F be a particular moral property and let G^* be a conjunction of natural properties. Moorean realism then involves the claim that there is no F and no G^* such that it is conceptually necessary that if x is G^* , then it is F . Neither is there an F and a G^* such that it is conceptually necessary that if x is G^* , then it is not F . This means that Moorean realism rejects the following two theses of strict entailment (where the modal operator denotes conceptual necessity):

$$(i) \quad \Box (\forall x)(G^*x \rightarrow Fx)$$

$$(ii) \quad \Box (\forall x)(G^*x \rightarrow \neg Fx)$$

Rejection of (i) entails acceptance of:

$$(i^*) \quad \Diamond (\exists x)(G^*x \ \& \ \neg Fx)$$

Rejection of (ii) entails acceptance of:

$$(ii^*) \quad \Diamond (\exists x)(G^*x \ \& \ Fx)$$

According to Blackburn, any coherent metaethical theory ought to be able to account for moral supervenience. Moorean realism ought therefore to be able to account for it. Blackburn contends, however, that Moorean realism is hugely problematical

because it does not facilitate such an account.²⁸ Where F is a particular moral property and G* is a conjunction of natural properties, and the modal operator denotes conceptual necessity, Blackburn's definition of moral supervenience is this:

F supervenes upon G* iff:

$$(S) \quad \Box [(\exists x)(G^*x \ \& \ Fx \ \& \ (G^*x \ U \ Fx)) \rightarrow (\forall y)(G^*y \rightarrow Fy)]^{29}$$

As I tried to argue in the previous chapter, there is reason to regard (S) as obscure. In an earlier version of the argument, Blackburn defines moral supervenience like this:

F supervenes upon G* iff:

$$(S^*) \quad \Box [(\exists x)(G^*x \ \& \ Fx) \rightarrow (\forall y)(G^*y \rightarrow Fy)]^{30}$$

This formula does not have the controversial U-operator, and is therefore much clearer than (S). This is not to say that (S*) is without problems, but it is at least clear what it says.³¹ Because (S*) is clearer than (S), I think that the strongest version of

²⁸(Blackburn 1973; 1984; 1985a).

²⁹(Blackburn 1985a:131)

³⁰(Blackburn 1984:184)

³¹ One problem is this: (S*) says that a certain conditional is necessarily true. The antecedent of the conditional is an existentially quantified sentence. This means that the antecedent is true if and only if there is something that satisfies the predicates in the existentially quantified sentence. If it is impossible that anything should satisfy these predicates, then it is impossible that the antecedent of the conditional should be true, and it is then impossible that the conditional should be false. If it is impossible that the conditional should be false, then it is necessarily true. Consider now a case in which 'G*' denotes a property that it is impossible that any object should instantiate. Under this interpretation of 'G*', it is impossible that the antecedent of the conditional should be true, and the conditional is therefore necessarily true. This means that F weakly supervenes upon G*. So, it is a result of using (S*) in a definition of weak supervenience that it can be shown that, for example, the property of being, unjust, (F) weakly supervenes upon the property of not being numerically self-identical (G*), or alternatively, upon the conjunctive property of distributing wealth according to the maxi-min principle and not distributing wealth according to the maxi-min principle. Actually, every moral property, weakly supervenes on any of these properties.

Another problem is this: the antecedent of the conditional will be vacuously true if it is impossible that something should simultaneously instantiate both F and G*. F and G* might be properties such that it is possible that numerically different objects, or the same object at different times, should instantiate one (but not both) of them. So, if everything that is meant by weak supervenience is captured by (S*), then it is correct to say that, for example, being unjust (F) weakly supervenes upon being completely made out of cement (G*). A third problem is that (S*) is vacuously true in cases where F and G* are properties instantiated by a single object, and it is impossible that any other object should instantiate these properties. This is the problem I touched upon in the previous chapter in relation to Kim's definition of weak supervenience.

Blackburn's argument involves reference to (S^*) and not (S). In my outline and discussion of Blackburn's argument, I shall therefore take it to involve (S^*).

By accepting (i^*), Moorean realism accepts that there is a conceptually possible world in which there is an object that is G^* and not F . By accepting (ii^*), it accepts that there is a conceptually possible world in which there is an object that is G^* and F . The truth of (S^*) excludes, however, that there is a conceptually possible world in which something is G^* and F , and something else is G^* and not F . But it seems odd that such a mixed world should be conceptually impossible when there is a conceptually possible world in which ' $(\exists x)(G^*x \ \& \ \neg Fx)$ ' is true, and a conceptually possible world in which ' $(\exists x)(G^*x \ \& \ Fx)$ ' is true. If it is conceptually possible that two such worlds should exist separately, why is it then, Blackburn asks, conceptually impossible that the two worlds should merge, and become a mixed world? In light of the truth of (i^*) and (ii^*), the truth of (S^*) seems mysterious.

It is important to note that this problem is not a formal one. The set of sentences made up by ' (S^*) ', ' (i^*) ' and ' (ii^*) ' is logically consistent, so by accepting the set, one does not violate any laws of logic. The problem is more of an informal kind: it poses an explanatory challenge to anyone who accepts this combination of sentences.

To ask why it is conceptually impossible that there should be a mixed world is equivalent to asking why (S^*) is true. There are, at least, two straightforward answers to this question. The first answer is that it is true of every object in every conceptually possible world, that if this object is G^* , then it is F . There are, in other words, no conceptually possible mixed worlds simply because there is no conceptually possible world in which something is G^* and not F . So, basically the

first answer is that (i) is true.³² Moorean realism cannot, however, give this as an answer since this position essentially involves the claim that (i) is false.

The second answer involves the claim that there is no conceptually possible world in which something is G^* and F . This is basically the answer that (ii*) is false. This answer is also unavailable to the Moorean realist. The reason for this simply is that his position involves the claim that (ii*) is true. But then the Moorean realist is left with no credible explanation as to why (S^*) is true. F 's supervenience upon G^* , in the way specified by (S^*), is underpinned by nothing. It is a brute conceptual fact that cannot be given any deeper explanation. Blackburn does not find this a philosophically credible view to be committed to.³³

Before addressing the question of how the Moorean realist might respond to Blackburn's argument, I want to clarify the argument in a certain respect. In his presentation of the argument, Blackburn suggests that acceptance of (S^*) and rejection of (i) is a necessary and sufficient condition for the argument to get going (Blackburn 1984:184; 1985a:137). However, this suggestion is not correct.

' $\neg \Box (\forall x)(G^*x \rightarrow Fx) \vdash_K \Diamond (\exists x)(G^*x \ \& \ \neg Fx)$ ' is valid. So by rejection of (i), one is committed to accepting that there is a conceptually possible world in which ' $(\exists x)(G^*x \ \& \ \neg Fx)$ ' is true. Rejection of (i) does not, however, entail that there is a conceptually possible world in which ' $(\exists x)(G^*x \ \& \ Fx)$ ' is true. ' $\neg \Box (\forall x)(G^*x \rightarrow Fx) \vdash_{SS} \Diamond (\exists x)(G^*x \ \& \ Fx)$ ' is invalid, and since the inference is invalid in the system with

³² This is the answer that an analytic naturalist would give. Analytic naturalism aims at establishing that for every moral property, F , there is a set of naturalistic properties, G^* , such that it is conceptually necessary that if something is G^* , then it is F . Whether or not the truth of analytic naturalism, and thereby the truth of (i) [and (ii)], can be established is a large and complicated question that for present purposes can be left to one side. For the moment, I just want to mention, and then shelve the possibility of countering Blackburn's argument by accepting (i) [and (ii)]. The important thing right now is to assess what damage Blackburn's argument does to a realist position that rejects (i) and (ii).

³³ In Blackburn's own words: "Supervenience becomes, for the [Moorean] realist, an opaque, isolated, logical fact for which no explanation can be proffered" (Blackburn 1973:119).

the greatest deductive powers, it is invalid in all the standard systems of modal logic.³⁴ Contrary to what Blackburn seems to think, no real mystery emerges from acceptance of (S*) and rejection of (i). This is so because one can explain the truth of (S*) by appealing to the truth of (ii). Acceptance of (S*) and (ii), together with rejection of (i) is logically consistent, and the truth of (ii) explains the truth of (S*). This would, however, be an odd combination of views to hold. If one rejects (i), what is the reason for accepting (ii)? A classical reason for rejecting (i) is that it is never the case that an object's instantiation of certain natural properties conceptually entails its instantiation of a certain moral property. If this is the reason for rejecting (i), then it is inconsistent not to reject (ii). So, strictly speaking, the argument only gets going against a realist position that accepts (S*) and rejects both (i) and (ii).

3. Responses to the Argument

The Moorean realist might respond to the argument by rejecting the claim that it is unacceptable to accept a supervenience relation as a brute conceptual fact. If it is acceptable to accept a supervenience relation as a brute conceptual fact, then it is of course no objection to Moorean realism that it offers no explanation of this fact. I do not think, however, that this response from the Moorean realist is satisfying. On the Moorean picture, moral concepts are used to pick out distinctive moral properties. Given this account of the semantic role of moral concepts, and given that it is conceptually possible that something G* should be F, and conceptually possible that something G* should be not F, it does in my opinion seem rather odd that it is conceptually impossible that something should be G* and F, and something else (in

³⁴A counterexample to the inference is: $\langle \{W_0, W_1\} ; \{DW_0 = a \in \neg G, F. DW_1 = a \in G, \neg F\} \rangle$.

that world) should be G* and not F. What explains this conceptual ban on mixed worlds?

A second response consists in saying that though a supervenience relation holds between the natural and moral properties of an object, it is not (S*) that should be used in a definition of this supervenience relation. The reason for this is that though moral supervenience is best captured by a thesis of weak supervenience, this relation does not hold with conceptual necessity, and if the modal operator of (S*) does not denote the same kind of necessity as the modal operator involved in (i*) and (ii*), the proposed argument against Moorean realism collapses.

Let me comment on this escape route for the Moorean realist. For reasons I rehearsed in the previous chapter, I think it should be accepted that moral supervenience holds with conceptual necessity. A distinctively conceptual mistake is made if one suggests that it is possible that two objects within a given world should differ in terms of being F without differing in terms of being G*.

Kim ascribes to Moore the view that supervenience is a fundamental synthetic, apriori fact that is not susceptible to further explanation (Kim 2000:13). Two things speak, in my opinion, against accepting Moore's view. Firstly, it simply gets the modality of moral supervenience wrong. Secondly, if Moore says that supervenience is a synthetic, apriori fact that cannot be given any deeper explanation, then he just postulates that a certain relation holds. To accept this is, in my opinion, to indulge in metaphysical extravagance. What is it that secures that identical moral properties are always instantiated in objects that are identical in terms of natural properties? Why is it metaphysically necessary that if one object is G* and F, then all G*-objects are F? Is it God who somehow secures that moral properties are always instantiated evenly across naturalistic similarities? If it is, then we have a situation in

which God could have distributed moral properties in objects in any way he liked, completely independently of the naturalistic properties of objects. For some strange reason he, however, chose to distribute them in exactly such a way that the world respects supervenience. I believe that we can safely dismiss this possibility. The Moorean realist must therefore come up with some other explanation as to why it is metaphysically impossible that naturalistically identical objects should have non-identical moral properties.

The Moorean realist might now suggest that, at least for some F's and G*'s, $(\forall x)(G^*x \rightarrow Fx)$ is necessary and synthetic. Where the modal operators denote metaphysical necessity, the truth of ' $\Box(\forall x)(G^*x \rightarrow Fx)$ ' explains the truth of ' $\Box[(\exists x)(G^*x \& Fx) \rightarrow (\forall y)(G^*y \rightarrow Fy)]$ '. This seems to be what Moore actually suggested. He writes:

"I should never have thought of suggesting that goodness was "non-natural", unless I had supposed that it was "derivative" in the sense that, whenever a thing is good (in the sense in question) its goodness (in Mr. Broad's words) "depends on the presence of certain non-ethical characteristics" possessed by the thing in question. I have always supposed that it did so "depend", in the sense that, if a thing is good (in my sense), then that it is so *follows* from the fact it possesses certain natural intrinsic properties, which are such that from the fact that it is good it does *not* follow conversely that it has those properties" (Moore 1942:588).

If an object's being good (being F) 'follows' from the fact that it instantiates certain natural properties (G*), then ' $\Box(\forall x)(G^*x \rightarrow Fx)$ ' must be true on Moore's picture. It simply makes no sense to say that being F 'follows' from being G* if it is possible that something should be G* and not F. The Moorean realist might now add that the

status of $(\forall x)(G^*x \rightarrow Fx)$ as being necessary and synthetic is not susceptible to any further explanation. The necessary connection expressed by the proposition is metaphysically and explanatory rock-bottom. To accept this view is, however, in my opinion, no less to indulge in metaphysical extravagance. Is it God who secures, for all objects and for all metaphysically possible worlds, that if an object is G^* , then it is F (where F , we must not forget, is a property from a class of properties distinct from the class of properties that G^* belongs to)?

Perhaps the Moorean realist replies to this by saying that asking this question is to ask for something that has no explanation and, importantly, does not need any explanation. If this is what the Moorean realist says, then I do not think that anything more can be said against him. We have arrived at a state of the debate where we are left with nothing but clashing intuitions about what kind of facts and relations are metaphysically unproblematic. I side with Hare when he writes:

"He [the moral realist] can, if we are prepared to swallow it, go on saying, without offence to supervenience, that there just is the sui generis non-natural property which all things of a certain kind necessarily (but not analytically) have...Like Blackburn, I find this hard to swallow, but others may have stronger stomachs (stomachs these days are getting harder)" (Hare 1984:7).

In general, I am of the opinion that Blackburn's argument is quite powerful. Its problems are in the technical department, and relate to the question of what formula should be used in a definition of the weak supervenience thesis that captures moral supervenience. None of Blackburn's two suggestions as to what formula should be involved in such a definition, are entirely satisfying. However, if Blackburn's argument is run with the definition of moral supervenience that I suggested in the previous chapter, then I think the argument retains its polemic force whereas its

vulnerability to technical objections decreases significantly. Blackburn's argument is powerful because it brings out clearly the point that Moorean realism has a problem with explaining moral supervenience when this relation is taken to hold with conceptual necessity. In addition to this problem, Moorean realism has, as I have tried to argue, a problem even if it retreats to the view that moral supervenience holds with metaphysical necessity. All in all, I therefore think that it should be concluded that Moorean realism fares badly on the issue of moral supervenience.

It must be stressed that even if it is true, as I claim it is, that Blackburn's supervenience argument is a powerful argument against Moorean realism, it is still a possibility that the argument is not a good argument against moral realism as such. There might be versions of moral realism that have no particular problem in accounting for the fact that moral supervenience holds with conceptual necessity. The question of whether or not Blackburn's argument is a good argument against moral realism as such, is a question that can only be answered after a careful discussion of how other versions of moral realism fare in comparison to Blackburn's argument. Until such a discussion has been carried out, one would be unwarranted in concluding that Blackburn's argument has any general scope.

In connection with this, it seems appropriate to mention that as it stands, the argument cannot function as an argument against analytic naturalism. It is a premise of Blackburn's argument that there is no F, and no G* such that (i) and (ii) are true. If this premise is true, then analytic naturalism is false. But this has nothing especially to do with supervenience. It is just a result of the fact that the truth of analytic naturalism is incompatible with the truth of this premise.

4. *Expressivism*

Blackburn has in a number of writings expounded and defended a position in metaethics that he calls 'quasi-realism'.³⁵ Central to this position is an expressivist analysis of moral utterances. The core idea of expressivism is that moral utterances have the function of expressing attitudes of approval or disapproval. A moral utterance such as 'x is wrong' does not express any belief of the speaker about x, and it does not primarily describe x. It is not semantically equivalent to an assertion that x possesses the property of wrongness. The utterance expresses, on the contrary, an attitude of disapproval towards x. Moral utterances do not have ordinary truth-conditions in the sense that certain states of affairs must obtain in order for the utterances to be true. This in turn means that the meaning of moral utterances is not exhausted by their truth-conditions. Blackburn has, relatively recently, given this description of what expressivism is:

"Expressivism denies that when we assert values [say such things as 'x is right' or 'x is just'], we talk about our own states of mind in actual or potential circumstances. It says that we voice our states of mind but denies that we thereby describe them" (Blackburn 1998:50).

Why is Blackburn eager to draw the distinction between voicing and describing one's states of mind and to stress that on the expressivist view, one does the former when one asserts one's values? An answer to this proceeds best via a short outline of some of the descriptivist analyses of moral utterances that Blackburn is opposed to.³⁶

According to descriptivism, moral utterances primarily describe the object that they are about. Both classical analytical naturalism, i.e. naturalism of the kind that Moore

³⁵See (Blackburn 1984; 1993a; 1998).

³⁶See (Blackburn 1984:167-171).

attempted to refute with his open question argument, and Moorean realism subscribe to a descriptivist analysis of moral utterances.

It makes sense to view the former position as a version of reductionism since it aims at reducing moral properties to natural properties, and tries to carry out this project by finding naturalistic synonyms for moral predicates. The underlying semantic idea is that predicates, in general, have a referring semantic function, and that synonymous predicates have the same referent. It is therefore the case, on this line of thought, that if it can be shown that some predicate, 'N', from the class of naturalistic predicates is synonymous with some predicate, 'M', from the class of moral predicates, then the predicates have the same referent, and we are therefore entitled to conclude that the moral property of M-ness is identical to N-ness. The moral utterance that 'x is M' is a description of x. The utterance is semantically equivalent to the assertion that x possesses a certain property: namely M. Moreover, the utterance has genuine truth-conditions in the sense that it is true if and only if x is N. If not the entire meaning of the utterance is exhausted by its truth-conditions, then at least its primary meaning is. One knows the primary meaning of the utterance when one knows the conditions under which it is true. Let us call this version of analytical naturalism, 'objective naturalism'.³⁷

According to non-naturalistic moral realism, moral utterances are descriptive of a special moral aspect of reality. Sui generis moral properties are, to quote Mackie, "part of the fabric of the world", and moral predicates refer to these properties.

³⁷There is a number of doctrines that can properly be said to be versions of analytic naturalism even though they depart somewhat from the general position outlined above. A proper generic name for these other doctrines might be 'subjective naturalism'. What all versions of subjective naturalism have in common is that they offer an analysis of moral predicates that involves reference to mental states. '1. person subjectivism' offers an analysis that involves reference to the mental states of the speaker who applies the predicate. 'Communitarian subjectivism' offers an analysis that involves reference to the mental states of the limited group of people who make up the community, society, organisation, tribe etc. to which the speaker is a member. Finally, 'universal subjectivism' offers an analysis of moral predicates that involves reference to the mental states of all people.

Saying that x is good is to give a description of x. It is to make the assertion that x has the property of goodness, and the utterance is true if and only if x has that property.³⁸

Blackburn dismisses the descriptive analysis of moral utterances put forward by objective naturalism on the ground that such an analysis makes it impossible to explain how two (or more) agents, that give different analyses of given moral predicate, can be involved in genuine moral disagreement if they differ in their views as to whether or not the predicate should be applied to a particular object. The argument is this: consider an agent who gives an analysis of the moral predicate 'M' ('right') to the effect that it means 'N' ('maximising overall preference-satisfaction'). Moreover, he analyses 'wrong' to mean 'not N' ('not maximising overall preference-satisfaction'). Consider now another agent that gives an analysis of 'M' to the effect that it means 'P' ('in accordance with God's commands'). On this second agent's analysis, 'wrong' means 'not P' ('not in accordance with God's commands'). Imagine now a situation in which the former agent says that some act is right whereas the latter agent says that it is wrong. Blackburn says that in such a situation, our intuitive response would be that the two agents are involved in a genuine moral disagreement. What the former agent says is, or so it seems, denied by the latter agent and vice versa. The fact that the two agents give different analyses of the moral predicates that they use has, however, the consequence that there in fact is no genuine disagreement between them. When the former agent says that the act is right, what he means is that it is maximising overall preference-satisfaction. He asserts, that is, that the object has this specific property.

³⁸(Mackie1977:15).

For there to be a genuine disagreement between him and the latter agent what has to be the case is that what the latter agent means by his claim that the act is wrong, is that the act is not maximising overall preference-satisfaction. This is, however, not what the latter agent means by his claim. What he means by his claim is simply that the act is not in accordance with God's commands. What we have is therefore a situation in which the former agent says that the act is maximising overall preference-satisfaction and the latter agent says that it is not in accordance with God's commands. This means, contrary to how things appear to be, that there is no genuine disagreement between the two agents. They do not contradict each other, and it is not the case that the beliefs expressed by the two agents cannot both be true.³⁹

Blackburn dismisses the descriptivist analysis put forward by Moorean realism on the ground that it presupposes indefensible ontological and epistemological commitments. He is also of the opinion that expressivism avoids the problems that various descriptivist analyses, according to him, have. Expressivism can account for moral disagreement:

³⁹In his discussion of analytic naturalism, Blackburn only considers objective naturalism (Blackburn 1984:168). The reason as to why Blackburn only considers objective naturalism is perhaps that he is of the opinion that the argument he offers against this version of analytic naturalism has broad scope in the sense that it refutes not only the considered version but all versions of analytic naturalism. To the extent that this is a correct interpretation of Blackburn's argumentative strategy, there is room for doubting that it achieves what it is supposed to achieve. Consider the analysis of moral utterances put forward by universal subjectivism. If one agent says that a specific act is right, then he makes the assertion that it is approved of by everybody. If another agent says that the act is wrong, then he makes the assertion that it is disapproved of by everybody. In this situation there is a genuine disagreement between the two agents. One might reply to this by saying that this descriptivist analysis of moral utterances is unsatisfactory because it gives a distorted account of what moral disagreement consists in. It construes moral disagreement as a disagreement about sociological facts, and it cannot possibly be the case, one might contend, that a moral disagreement between two agents is a disagreement about such matters. Two things should be said in response to this: firstly, whether or not this analysis of moral utterances misconstrues the nature of moral disagreement is a completely different issue from what is raised by Blackburn's argument. If it misconstrues the nature of moral disagreement, then this is something that must be shown and not just assumed and, importantly, it takes additional argument to what is provided by Blackburn to show this. Secondly, if the above objection is grounded in the conviction that a proper analysis of moral utterances construes moral disagreement as a disagreement in attitude and not a disagreement about facts (irrespective of these facts being sociological, economical or something else), then one simply begs the question against a descriptivist analysis of moral utterances.

"The expressive theory avoids this undesirable consequence. It locates the disagreement where it should be, in the clash of attitudes towards [the act]" (Blackburn 1984:168).

When one agent says that some act is right and another agent says that it is not right (or says that it is wrong), then there really is a genuine moral disagreement between them. There is such a disagreement because what is happening in the situation is that the two agents express opposing attitudes towards the act. The former agent expresses an attitude of approval towards the act. By saying that the act is not right, the latter agent expresses a different attitude towards the act. It is not a distinct attitude of disapproval towards the act, but it is nonetheless an attitude that opposes and clashes with the attitude expressed by the former agent.⁴⁰

By denying that moral predicates refer to *sui generis* properties, expressivism avoids problems of accounting for how these properties fit into a general naturalistic world view, and how we have epistemological access to them.

What should we think of Blackburn's arguments concerning what analysis to give of moral utterances? In answering this question, one should separate Blackburn's arguments into two classes: the negative arguments against the various forms of descriptivism, and the positive arguments for the adequacy of an expressivist analysis of moral utterances. For present purposes, no stand needs to be taken on the soundness of Blackburn's negative arguments. A stand needs, however, to be taken on his positive ones. Expressivism presents us with a simple and coherent account of what moral disagreement consists in. Its ontological and epistemological

⁴⁰To see that the two attitudes actually clash, imagine them being attitudes of a single agent. Such an agent would have an inconsistent set of attitudes. The inconsistency of this is clearly brought into the open by considering the fact that there is no possible world that conforms to the pattern of attitudes expressed by this agent. By approving of *x* (the act), the agent expresses a view to the effect that *x* is realised in every (morally ideal) possible world. By not approving of *x*, the agent expresses a view to the effect that there is a (morally ideal) possible world in which *x* is not realised. It is, however, logically impossible that there should be a world in which both *x* and not *x* is realised.

presuppositions are, moreover, minimal. These features of the theory are, in my opinion, enough to establish expressivism as an interesting semantic theory that is worthy of further attention. If expressivism can proffer a convincing explanation of moral supervenience, then its credentials are further enriched. So, the question that needs to be addressed now is the dual one of what the expressivist explanation of moral supervenience is, and whether or not it is convincing.

5. The Expressivist Explanation of Moral Supervenience

Expressivism denies theses (i) and (ii) presented in section two. Since expressivism and Moorean realism are allied in the rejection of these theses, and since the latter's rejection of these theses are, at least partially, responsible for the fact that it faces a problem in relation to moral supervenience, one might wonder if expressivism will not face the same problem. Blackburn is, however, of the opinion that the rejection of (i) and (ii) yields no problem for expressivism in connection with moral supervenience. This theory is quite capable of giving an explanation of moral supervenience that can accommodate the fact that it holds with conceptual necessity. The explanation is this:

"From the anti-realist point of view things are a little easier. When we announce the A-commitments [moral judgements] we are projecting, we are neither reacting to a given distribution of A-properties [moral properties], nor speculating about one. So the supervenience can be explained in terms of the constraints upon proper projection. Our purpose in projecting value predicates may demand that we respect supervenience. If we allowed ourselves a system (shmoralizing) which was like ordinary evaluative practice, but subject to no such constraint, then it would allow us to treat naturally identical cases in

morally different ways. This could be good schmoralizing. But that would unfit schmoralizing from being any kind of guide to practical decision-making (a thing could be properly deemed shbetter than another although it shared with it all the features relevant to choice or desirability" (Blackburn 1984:186).

Supervenience is here explained as a conceptual constraint upon what it is to be 'moralising'. If one does not respect supervenience and applies non-identical evaluative predicates to objects that one judges to be naturalistically identical, then the practice that one is engaged in is simply not the practice of moralising. It is some other practice that we might call 'schmoralizing'. Why cannot an evaluative practice that is not disciplined by supervenience count as moralising? This has something to do with the distinctive purpose of the practice of moralising. The practice of moralising serves, on the expressivist picture, the purpose of being a guide to practical decision-making. To apply an evaluative predicate to an object is not only to express an attitude of approval or disapproval towards it. It is also to commend or condemn it, and to rank it either favourably or disfavourably in comparison to other objects. The practice of moralising can only serve its purpose if it is disciplined by respect for supervenience.

If an individual could be said to be moralising while adopting an evaluative practice that consists in applying non-identical evaluative predicates to objects that are naturalistically identical, then the practice of moralising could not function as a guide to practical decision-making. Other people looking to the evaluative practice of this individual for guidance concerning what to do, desire and choose, would be left in a state of confusion since the individual would express inconsistent attitudes and recommendations towards objects that are naturalistically identical. On some

occasions he would approve of, commend and rank favourably objects with natural properties N, and on other occasions, he would disapprove of and rank disfavouredly such objects.

To make the expressivist account of moral supervenience even more clear, consider this passage:

"It seems to be a conceptual matter that moral claims supervene upon natural ones. Anyone failing to realize this, or to obey the constraint, would indeed lack something constitutive of competence in the moral practice. And there is good reason for this: it would betray the whole purpose for which we moralize, which is to choose, commend, rank, approve, or forbid things on the basis of their natural properties" (Blackburn 1985a:137).⁴¹

Again, the view is that it is a conceptual constraint upon using moral vocabulary that one respects supervenience and applies identical moral concepts to objects that one judges to be naturalistically identical. If one does not respect supervenience in one's use of moral concepts, then one displays conceptual confusion and makes evident that one is not a competent user of moral concepts. It is, according to Blackburn, simply constitutive of competence in the moral practice that one respects supervenience. Two people can use moral concepts in obedience to the conceptual constraints that govern all use of moral concepts even though they adopt different standards, and do not apply the same moral concept to a given object characterised by a certain set of natural properties. Both trespass, however, on the conceptual constraint that govern all use of moral concepts if they do not respect supervenience.

⁴¹One thing to note here is that Blackburn implicitly rejects the view that supervenience in the realm of ethics is a relation between properties. He commits himself to ascriptive supervenience by saying that it is moral claims (or judgements) that supervene upon natural claims (or judgements).

What is the explanation as to why it is constitutive of competence with the moral vocabulary that supervenience is respected? Again, the explanation has as its starting point a certain view as to what the purpose is of using moral vocabulary. As the quoted passage attests, it is Blackburn's view that the purpose for which we use moral vocabulary is to commend, rank approve and disapprove of things because of their natural properties. Put slightly differently, the role of using moral vocabulary is to guide choices and desires among the natural features of the world. Its role is not to describe or depict a distinct moral aspect of reality. Now, our use of moral vocabulary can only fulfil this purpose or play this role if it is disciplined by respect for supervenience. An evaluative practice in which it was acceptable to break supervenience and apply non-identical moral concepts to objects that are judged to be naturalistically identical would simply defeat or betray the purpose for which we use moral vocabulary.

If one accepts the expressivist assumptions about the semantic role of moral judgements and the purpose for which we apply moral concepts to objects, then it appears to me that expressivism has a coherent and convincing explanation as to why it is constitutive of competence with the moral vocabulary that the supervenience constraint is respected. I therefore do not think that there is anything in Blackburn's explanation of moral supervenience that should make us believe that the expressivist has an independent problem on the issue of moral supervenience. Expressivism fares well on this issue.

Of course, one could question and argue against the expressivist assumptions about the semantic role of moral judgements, and the purpose for which we apply moral concepts. This line of critique would not, however, be a direct critique of the expressivist explanation of moral supervenience. It would be a critique of the

expressivist assumptions on which Blackburn builds his explanation of moral supervenience. Such a critique may or may not be justified, but since the question under consideration here is the question of whether or not the expressivist has an independent problem on the issue of moral supervenience, it is not an interesting line of critique to attack the key assumptions of expressivism. This is not to be overgenerous towards expressivism, or in any other way to be playing it into the hands of expressivism. It is simply to be seeking an answer to the question of whether or not expressivism can account for moral supervenience.

Chapter 5

Round One of the Frege-Geach Point

1. *Introductory*

In this chapter, I shall continue the exploration of expressivism. The focus will be directly on expressivism's core thesis, and I shall start answering the question of whether or not expressivism, in the end, is a plausible semantic theory. What I intend to do is to consider in detail the objection to expressivism that historically has proved to be the most serious and persistent. The objection in question is what has come to be known as the 'Frege-Geach point'. At this point, I will only be concerned with the initial stage of the debate relating to the Frege-Geach point. The content of the chapter is therefore mainly exegetical. A chapter of this nature is, however, helpful for the understanding of the more recent developments on the issue that I shall be concerned with in the two subsequent chapters.

In section two, I give an account of what the Frege-Geach point is, and why it might be thought to be a problem for a proponent of expressivism. Section three contains an outline of Blackburn's response to the problem posed by the Frege-Geach point. This response involves, among other things, the development of a formal language into which sentences in natural language can be translated. I give an account of this formal language, and discuss some of its features. In sections four and five, I critically discuss Blackburn's response. First, I argue for a certain way of understanding what attitudinal inconsistency consists in. I then consider points made by Hale and Wright, and I end by concluding that Blackburn's strongest defence is

defective. This is so because it does not allow us to say that a distinctively logical mistake is made if one accepts the premises but not the conclusion of a modus ponens inference that involves sentences that are given an expressivist analysis.

2. *The Problem*

The recent genesis of the Frege-Geach point is in an article by Peter Geach.⁴² With the exception of one quote from the original source, I shall, however, take Blackburn's presentation and discussion of the issue in *Spreading the Word* as my starting point.⁴³ At the beginning of his article, Geach observes that:

"a proposition may occur in discourse now asserted, and now unasserted, and yet be recognizably the same proposition".

Consider the sentence, 'It is wrong to tell lies'. This sentence contains the predicate 'wrong' and the sentence can occur both asserted and unasserted. The sentence occurs asserted when it occurs on its own and it occurs unasserted, or embedded, when it occurs as a constituent sentence of a larger sentence such as a conditional or disjunction.⁴⁴

An example of a context in which the sentence occurs unasserted is the conditional 'If it is wrong to tell lies, then it is wrong to get your little brother to tell lies'. By affirming this sentence, one does not affirm the sentence 'It is wrong to tell lies'. One affirms a conditional of which the original sentence is the unasserted antecedent. The Frege-Geach point is that a sentence such as 'It is wrong to tell lies' means the same whether or not it occurs asserted or unasserted. In other words, the

⁴² See (Geach 1964).

⁴³ (pp:189-196)

⁴⁴ The sentence can occur unasserted in other contexts as well: contexts such as: 'It is not the case that it is wrong to tell lies' and 'He believes that it is wrong to tell lies'.

sentence does not change meaning in a transition from an asserted to an unasserted occurrence. The proof of this is that the following modus ponens inference is valid:

- 1) If it is wrong to tell lies, then it is wrong to get your little brother to tell lies.
- 2) It is wrong to tell lies.
- 3) It is wrong to get your little brother to tell lies.

A necessary condition for the validity of the argument is that the sentence 'It is wrong to tell lies' means the same in each of its occurrences in the argument. If the sentence does not have the same meaning in each of its occurrences, then there is a fallacy of equivocation.

Why should all this constitute a problem for expressivism? The problem is this: in 2), 'It is wrong to tell lies' occurs asserted and according to expressivism, an attitude of disapproval towards telling lies is expressed when the sentence is affirmed. This means that in 2), the sentence has a distinctive expressive meaning. In 1), the same sentence occurs in an unasserted context and by affirming 1), no attitude of disapproval towards telling lies is expressed. Somebody who affirms 1), affirms a conditional and cannot properly be said to be expressing an attitude of disapproval towards telling lies. This means that 'It is wrong to tell lies' does not have, as it occurs in 1), a distinctive expressive meaning. The validity of the modus ponens inference shows, however, that 'It is wrong to tell lies' means the same whether or not the sentence occurs asserted or in an unasserted context. It cannot therefore be the case that the sentence has a distinctive expressive meaning when affirmed on its own.⁴⁵ Expressivism is therefore wrong.

⁴⁵This conclusion can be resisted by denying that the inference is valid. If one simply denies that the inference is valid, then there is no inconsistency involved in holding the view that the sentence shifts meaning from an unasserted to an asserted context. Such a denial is, however, very costly. It involves rejecting central theses in classical logic and doing this has severe consequences for how one can

As a first comment to the Frege-Geach point, Blackburn notes that expressivism does not have to concede defeat before the possibilities have been explored of giving an account of unasserted contexts that is such that the evaluative constituent sentences of such contexts have a distinctive expressive meaning (Blackburn 1984:191). The underlying idea of this thought is that the availability of any such account will enable the expressivist to hold that an evaluative sentence has a meaning in unasserted contexts that is identical to the meaning it has when it is affirmed on its own. What must be included in such an account, according to Blackburn, is an explanation of what we are up to when we make use of linguistic expressions in which evaluative sentences occur unasserted. What is it, for example, that we are up to when we express ourselves in terms of conditionals with evaluative constituent sentences?⁴⁶

Understanding of Blackburn's answer to this question goes via understanding of the notion of a 'moral sensibility'. On the projective picture, a moral sensibility is a function from input of beliefs to output of attitude. Our moral psychology works, according to projectivism, in such a way that we form beliefs about how the world is in terms of natural properties and any moral conviction is then to be seen as an attitudinal response to the world. Two agents may agree about the natural properties of an act but disagree in their attitudinal response to the act.

The moral sensibility of the one agent may be such that the act's natural properties triggers an attitude of disapproval while the other agent has a moral sensibility that is such that the very same set of natural properties triggers no such

argue on philosophical issues unrelated to the Frege-Geach point.

⁴⁶ In *Spreading the Word*, Blackburn focuses solely on conditionals, but the Frege-Geach point has general scope. All unasserted contexts, not only conditionals, pose an initial problem for an expressivist analysis of moral utterances. This means that Blackburn, in order to make expressivism credible, must be able to give an account of each and every unasserted context and the account must be such that the evaluative sentence occurring unasserted preserves its distinctive expressive meaning.

attitude. According to Blackburn, not all moral sensibilities are admirable. Some are deplorable in the sense that they trigger an attitude of approval (or disapproval) towards acts and social institutions that, according to a specific first order moral view, do not deserve such an attitudinal response. Other sensibilities are unadmirable because they are unreliable in the sense that they on different occasions trigger attitudes that are inconsistent with each other.

Blackburn is of the opinion that it is very important to us to rank various sensibilities and to endorse some sensibilities and condemn and reject other sensibilities. Why is this important? It is important because people's behaviour is, to a large extent, a function of their moral sensibility and one of the things that have the most severe consequences on the desirability of the world we live in is the way other people behave. A world in which people keep their promises, speak the truth and engage in acts of co-operation is much more desirable than a world in which people do not do these things.

An important feature of any moral sensibility is the way it pairs together different first order attitudes. It is, for example, easy to think of a sensibility that pairs an attitude of disapproval towards telling lies with an attitude of approval towards getting other people to tell lies. If one's own moral sensibility is such that one finds such a pairing of individual first order attitudes repugnant, what linguistic expression should one make use of in order to express one's disapproval of this pairing?

Blackburn's suggestion is that the evaluative conditional is the natural way of expressing approval of specific pairings or combinations of first order attitudes. Somebody who affirms 'If it is wrong to tell lies, then it is wrong to get your little brother to tell lies' expresses an attitude of approval towards actually combining, or having a disposition to combine, an attitude of disapproval towards telling lies with a

similar attitude towards getting your little brother to tell lies. This means that the function of the conditional is to express a substantial moral view or opinion: "But it is quite satisfactory that the conditional expresses a moral point of view" (ibid.:193). People with the moral view that there is a big difference between telling lies yourself and getting your little brother (or anybody else, for that matter) to tell lies will reject or abstain from affirming the conditional. They will reject the conditional because they reject the moral view that an attitude of disapproval towards telling lies should be combined with a similar attitude towards getting other people to lie.

3. *The Formal Language E_{ex}*

This is Blackburn's general account of what we are up to when we express ourselves in terms of a conditional with evaluative constituent sentences. He now supplements this account with a semantic theory. We are invited to imagine a language that is very much like ordinary English. The only difference is that it does not contain any evaluative predicates such as 'good', 'right' and 'wrong'. The expressive nature of value judgements in this language, *E_{ex}*, is completely transparent because *E_{ex}*, contains a 'hooray!' operator and a 'boo!' operator (H!, B!) that attach to descriptions of things to give expressions of attitude. 'H!(telling the truth)' expresses approval of telling the truth and 'B!(telling lies)' expresses disapproval of telling lies.

For the reasons already developed, speakers of *E_{ex}* will want a notation for expressing attitudes towards various couplings of attitudes. Blackburn suggests that we talk about an attitude by putting its expression inside bars. The expression '/H!(X)/' does not express any attitude towards X but can be used to refer to approval of X. This means that if a speaker of *E_{ex}* wishes to express disapproval of approval of X, then he should make use of this expression: 'B!(/H!(X)/)'. Another syntactical

device of E_{ex} is the semicolon, ';'. The semicolon is used to denote the view that one attitude involves or is coupled with another.⁴⁷ With these technical devices in place, it is now possible to reconstruct how a speaker of E_{ex} will express himself if he is to express the moral view expressed by the evaluative conditional of ordinary English:

4) H! (/B!(telling lies);/B!(getting your little brother to tell lies)/).

4) expresses an attitude of approval of making (disapproval of getting your little brother to tell lies) follow upon (disapproval of telling lies). In this reading of 4), which is the one Blackburn favours, an expression of the form 'A;B' is not taken to denote the view that attitude A involves or is coupled with attitude B, but is interpreted as meaning 'making B follow upon A'. This means that Blackburn has three possible interpretations of the semicolon in play. 'A;B' can mean either of 'A involves B', 'A is coupled with B' or 'making B follow upon A'. There might not be much that separates the three readings but it seems to me that Blackburn's exposition of this syntactic device of E_{ex} would have benefited from a bit more clarity and stringency. It is quite reasonable for the reader to wonder at this point what, if anything, hangs on the shift in interpretation of the semicolon.

Blackburn suggests that E_{ex} will be spoken by people who need to signal and respect consistencies and inconsistencies. Consider now an agent who holds this pair of attitudes:

4) H! (/B!(telling lies);/B!(getting your little brother to tell lies)/).

5) B! (telling lies).

Is this agent committed to holding an attitude of disapproval towards getting your little brother to tell lies? Blackburn gives an affirmative answer:

⁴⁷"And suppose we use the semi-colon to denote the view that one attitude or belief involves or is coupled with another" (ibid.:194).

"Anyone holding this pair must hold the consequential disapproval: he is committed to disapproving of getting little brother to lie, for if he does not his attitudes clash" (ibid.:195).

The important thing here is the notion of attitudes that 'clash'. Blackburn's reason for thinking that the agent's attitudes clash is this: by holding the attitude expressed by 4), the agent has an attitude of approval towards making disapproval of getting your little brother to tell lies follow upon disapproval of telling lies. By holding the attitude expressed by 5), the agent has an attitude of disapproval towards telling lies. By failing to have an attitude of disapproval towards getting your little brother to tell lies, the agent's attitudes might be said to clash in the following sense: since the agent fails to have an attitude of disapproval towards getting your little brother to tell lies and actually has an attitude of disapproval towards telling lies, he fails to make disapproval of getting your little brother to tell lies follow upon disapproval of telling lies. He therefore fails to do something that he himself approves of.

Put differently, and using an alternative interpretation of the semicolon: by failing to have an attitude of disapproval towards getting your little brother to tell lies, the agent fails to couple an attitude of disapproval towards telling lies with an attitude of disapproval towards getting your little brother to tell lies, and he therefore fails to couple attitudes in a way that he himself approves of.

Blackburn goes on to say that anybody who holds the attitudes expressed by 4) and 5) but does not hold the 'consequential disapproval' has a fractured sensibility that cannot be an object of approval. The reason why such a sensibility cannot be an object of approval is that it cannot fulfil the practical purposes for which we evaluate things. As noted in the previous chapter, the expressivist picture has it that we evaluate things in order to guide behaviour and practical decision-making. To apply

the predicate 'good' or 'right' to an act is, among other things, to recommend to others the performance of the act. It is clear that if an agent's moral sensibility is fickle in the sense that it issues in attitudes that clash with each other (attitudes that are internally inconsistent), then the agent's practice of evaluating things cannot fulfil its practical purposes. The agent will in his evaluative practice both encourage the performance and the non-performance of certain acts and such a practice cannot function as a guide to behaviour and practical decision-making.

Blackburn is of the opinion that E_{ex} gives the 'deep' or 'logical' structure of our evaluative remarks. Evaluative remarks have a propositional surface-form, but the logical structure of any such remark can be represented in E_{ex} . This does not mean, however, that there is something mistaken in putting our evaluative commitments in propositional form and applying the truth-predicate to evaluative remarks. Since the desirability of the world we live in to a large extent depends upon what attitudes other people hold, it is important to us that we are able to express concern for the consistency of other people's attitudes and are able to express agreement and disagreement with their attitudinal stance.

We are, no doubt, able to do these things by speaking E_{ex} but our practical need to debate, ponder and express agreement and disagreement with expressions of attitudes might be easier dealt with if we invent predicates answering to the two kinds of attitude and instead of using long and unhandy expressions such as 'B! (telling lies)', 'H! (keeping one's own promises)' and 'H! (/B!(telling lies)/;B!(getting your little brother to tell lies)/)' we simply express ourselves in propositional form as in 'It is wrong to tell lies', 'It is good to keep one's own promises' and 'If it is wrong to tell lies, then it is wrong to get your little brother to tell lies'.

Since all evaluative remarks in ordinary English can be represented in E_{ex} , it is possible to reconstruct the original modus ponens inference in E_{ex} . It looks like this:

- 4) H! (/B!(telling lies)/;/B!(getting your little brother to tell lies)/).
- 5) B! (telling lies).
- 6) B! (getting your little brother to tell lies).

It is clear that Blackburn would commit a mistake if he says that the original inference is valid because it is impossible that the premises should be true and the conclusion false. As the reconstruction of the inference in E_{ex} shows, neither the premises nor the conclusion are sentences that express a proposition. Blackburn does not make this mistake. His thought is that what underpins the inference is the idea, already described, that affirmation of the premises and failure to accept the conclusion involve a clash of attitudes (ibid.:195).

4. *Critiques of Blackburn's Response*

My first point of critique of Blackburn's response to the Frege-Geach point centres on Blackburn's thought that one has clashing attitudes if one affirms 1) and 2) but fails to accept 3). It seems to me to be a misdescription of such an agent to say that his attitudes clash. To see this, we need to understand what the necessary and sufficient conditions are for a clash of attitudes. When we have to our disposal the two operators that Blackburn has introduced, it seems that a clash of attitudes can obtain in one of three ways: an agent has a set of clashing attitudes just in case he has at least one of the following pairs of attitudes: 'H! (X)' and 'B! (X)', 'H! (X)' and 'H! (not-X)' or 'B! (X)' and 'B! (not-X)'.⁴⁸ If this is accepted, then it is not true to say that

⁴⁸ This view about what conditions must obtain in order for attitudes to clash is very similar to that expressed by Hale: "But we might, I think, say that attitudes clash if they consist of approval and

the agent under consideration has a set of clashing attitudes. He has none of the three pairs of attitude. He has an attitude of disapproval towards telling lies and fails to have an attitude of disapproval towards getting your little brother to tell lies.

It is a mistake to infer from this failure that he has an attitude of approval towards getting your little brother to tell lies. There is logical space to occupy between approval and disapproval of X. If one does not disapprove of X, then one does not necessarily approve of X. One can have an attitude of toleration or indifference towards X. We do not know what the agent's reason is for not accepting 3). It can either be because he has an attitude of approval towards getting your little brother to tell lies, or because he has an attitude of toleration or indifference towards it.

Whichever of these two attitudes he actually holds, it is not the case that this attitude in combination with his attitude of disapproval towards telling lies clashes with his attitude of approval towards making disapproval of getting your little brother to tell lies follow upon disapproval of telling lies. Only two attitudes clash with the attitude expressed by 1): namely the ones expressed by:

7) B! (/B!(telling lies)/;/B!(getting your little brother to tell lies)/)

8) H! \neg (/B!(telling lies)/;/B!(getting your little brother to tell lies)/)⁴⁹

and the agent has neither of these. In my opinion, there is therefore reason to be sceptical about the coherence of the idea that the agent has clashing attitudes.

Blackburn is mistaken in saying that an agent affirming 1) and 2) is, on pain of having clashing attitudes, committed to accepting 3).⁵⁰

disapproval towards the same action-type, or if they are attitudes of approval towards x-ing and not-x-ing" (Hale 1986:74).

⁴⁹The negation involved in 8) is not the ordinary, truth-functional one. Informally, 8) should be read as expressing approval towards not combining the attitudes conjoined by the semicolon.

⁵⁰In saying this, I to some extent echo what Wright says in this passage: "Blackburn does indeed speak of the 'clash of attitudes' involved in endorsing the premises of the modus ponens example, construed as he construes it, but in failing to endorse the conclusion. But nothing worth regarding as

5. *An Alternative Analysis of 1)*

If there is no case for saying that the inference to 3) from 1) and 2) is underpinned by the fact that acceptance of 1) and 2) together with failure to accept 3) result in a clash of attitude, what is it then that underpins the inference? Bob Hale has, on behalf of the expressivist, offered an analysis of 1) that is different from the one given by Blackburn (Hale 1986:74). According to the alternative analysis, affirming the evaluative conditional commits one to disapproving of anyone who disapproves of telling lies but does not disapprove of getting your little brother to tell lies. This attitudinal commitment can only be represented in E_{ex} if E_{ex} is supplemented with a kind of negation ' \sim '. This sign enables us to represent 'lack of attitude'. ' $\sim/H!(X)!$ ' denotes, for example, lacking an attitude of approval of X. On the alternative analysis, 1) is therefore the surface-form of:

9) $B! (/B!(\text{telling lies}); \sim/B!(\text{getting your little brother to tell lies})/)$

Consider now an agent who has this attitudinal commitment and who moreover affirms 2) and fails to accept 3). An agent with this pattern of attitudes combines an attitude of disapproval towards telling lies with lack of disapproval towards getting your little brother to tell lies. By affirming 1), the agent, however, explicitly disapproves of such a combination of attitudes. If this alternative analysis is adopted, then the expressivist can say that an agent affirming 1) and 2) but fails to accept 3) is involved in a special kind of inconsistency.

The agent is not inconsistent in the sense that he has clashing attitudes but since it, according to Hale, makes sense to see an agent's adoption of an attitude as an action of his⁵¹, the agent is inconsistent in the sense that he fails to make his actions

inconsistency seems to be involved. Those who do that merely fail to have every combination of attitudes of which they themselves approve" (Wright 1988:33).

⁵¹This appears to be correct. It would be a mistake to think that what falls under the concept of an action is only a kind of intentional bodily movement that causes physical change to the environment in

cohere with his attitudinal commitments. The agent disapproves of any sensibility that combines an attitude of disapproval towards telling lies with toleration or condonation of getting your little brother to tell lies. This is a moral principle of his. He, however, fails to make his own actions cohere with this principle since he adopts the very combination of attitudes that he disapproves of.

Another way of describing the agent, one that goes beyond Hale's way of putting matters, is to say that the agent fails to live up to his own ideals or values. He sees avoidance of a certain combination of attitudes as an ideal or something desirable but he then fails to live up to his own ideal in the sense that he adopts the very combination of attitudes that he disapproves of. Hale suggests that the proposal sketched by him:

"appeals to a notion of moral inconsistency which is plausible independently of projectivist leanings - one is guilty of a morally important kind of failing (we might as well call it inconsistency) to the extent that one fails to make one's actions accord with one's principles" (ibid.:74).

It seems to me that Hale's alternative analysis of 1) together with the addition of the negation sign to E_{ex} marks an improvement upon Blackburn's original suggestions. The special merit of Hale's proposal is that it allows the expressivist the view that one is involved in inconsistency if one affirms the premises of the original inference but fails to accept its conclusion. The expressivist can then say that what underpins the original inference is the fact that acceptance of its premises combined with failure to accept its conclusion yields inconsistency in the explained sense.

which the agent is placed. This concept of an action would preclude the intelligibility of a notion such as an 'act of thinking' and it would make it wrong to suggest that when an agent tries to remember something, he is performing a certain act. If 'mental acts' such as thinking and remembering can properly be said to fall under the concept of an action, as I think they can, then I can see no reason why it should be wrong to say that an agent performs an action when adopting a (new) attitude.

It should be noted that Hale's suggestions do not constitute any major problem for Blackburn. In the broad perspective, the suggestions are minor corrections and additions to a general kind of analysis developed by Blackburn and no loss of standing will fall upon Blackburn if he simply accepts these minor corrections and additions. Blackburn can rightfully say that far from showing the utter impossibility of giving an expressivist analysis of evaluative conditionals, Hale's suggestions actually do something to bolster the idea that such an analysis is tenable.

Before concluding that Blackburn has, with the help of Hale, presented us with a convincing expressivist analysis of evaluative conditionals and thereby done much to show that unasserted contexts do not pose a problem for expressivism, we should, however, consider a rather classic objection put forward by Hale.⁵² The objection has its starting point in the question of what is wrong, on the expressivist account, with affirming the premises of an instance of a moral modus ponens and then failing to accept the conclusion of the inference? Put slightly differently: what kind of mistake is made by someone who accepts the premises but not the conclusion of such an inference?

The answer we get is that the mistake is of a moral kind. The agent who accepts the premises but fails to accept the conclusion seems to be morally flawed in the sense that he fails to be true to one of his own moral principles. He does what he himself explicitly disapproves of. It is possible, for the sake of the argument, to agree that this amounts to some sort of inconsistency on the part of the agent but, importantly, the expressivist account does not give us the resources to say that the agent commits a logical mistake or that he is logically confused in accepting the

⁵²Hale's objection is in (Hale 1993a:344). The objection owes much, as Hale himself recognises, to considerations made by Crispin Wright. See (Wright 1988).

premises but failing to accept the conclusion of the inference. Wright has, however, urged that there is a strong intuitive pull to regard the agent as committing a mistake that is not, or at least not merely, a moral mistake (Wright 1988:33).

I agree with this. It seems intuitively right to say that someone who affirms 1) and 2) but fails to accept 3) has, in addition to whatever other flaws he may possess, a flaw in his logical reasoning. The expressivist account does not at all square with this strong intuitive pull. The expressivist account, whether in its original or amended form, of an evaluative conditional, is therefore mistaken. This is a conclusion that Blackburn accepts. In a later work, he presents a new analysis of evaluative conditionals: an analysis that aims at steering clear of the difficulties that faced his previous one.

Chapter 6

Round Two of the Frege-Geach Point

1. *Introductory*

My main aim in this chapter is to give an account and critique of the formal logical system that Blackburn develops in the article "Attitudes and Contents" (AC).⁵³ The account of this system is found in sections two to five. In these sections, I try, among other things, to make clear how this new logic for expressions of attitude enables Blackburn, given certain assumptions about the notion of validity, to show that Geach's inference is valid. Moreover, I give an account of Blackburn's way of explaining the meaning of standard connectives in cases where these operate on sentences that are not truth-apt. Section six contains a discussion of one of the rules of Blackburn's logical system, and I argue that Blackburn fails in an attempt to overcome a dilemma set out by Hale. In section seven, I argue for the claim that Blackburn's assumption about the well-formedness of a certain formula in his formal language yields a significant problem that does not admit of any obvious solution. Section eight contains the two conclusions of the chapter. The first conclusion is that, because of the problems discussed in sections six and seven, Blackburn's logical system is unsatisfactory as it stands. The second conclusion is more positive in nature, and it is that Blackburn, with the introduction of a commitment-semantics for binary connectives, has suggested an interesting route forward for the expressivist: a route that deserves further and detailed attention.

⁵³(Blackburn 1988).

2. A New Formal Language

The problem that a conditional with evaluative constituent sentences presents for the expressivist is a particular instantiation of a more general problem. As previously mentioned, evaluative sentences can occur unasserted, not only in conditionals, but in a number of different contexts. Blackburn's aim in AC is, at least partly, to offer an analysis that has general scope in the sense that it can be used to explain and make sense of the character and status of evaluative sentences in all the unasserted contexts that such sentences can occur in. In AC, Blackburn abandons the view that E_{ex} is a formal language that can be used to bring out the underlying structure of our ordinary evaluative talk. The formal language that Blackburn introduces as a replacement for E_{ex} is supposed to exhibit explicitly the expressive nature of ordinary moral discourse, and it is a language that syntactically is very similar to the language of standard deontic logic. The two sentence-forming operators of standard deontic logic, 'O' and 'P' are, however, replaced with two attitude-operators: namely the familiar 'H!' and a new one, 'T!'.

In the new approach, the two sentence-forming operators are applied to sentence-letters. In this they are similar to the two sentence-forming operators of standard deontic logic. In deontic logic, the formula 'Op' is, when given an interpretation, a report that p is obligatory according to some background set of norms. The formula is therefore an expression with a genuine truth-condition. The two attitude-operators can, in contrast to this, be used to form formulae that are expressive of attitude. Applying 'H!' to a sentence letter results in a formula that, when given an interpretation, is expressive of attitude. According to Blackburn, 'H!p' "can be seen as expressing the view that p is to be a goal, to be realised in any perfect world" (Blackburn 1988:189).

Consider this sentence in natural language: 'It is right to keep promises'. On the expressivist picture, an agent who advances this sentence, expresses approval of keeping promises. Such an agent can be seen as expressing approval of a certain state of affairs obtaining: namely a state of affairs in which promises are kept. This state of affairs is represented by the sentence 'promises are kept'. The underlying form of the sentence is now seen to be 'H!p' where 'p' is the sentence 'promises are kept'. An agent who approves of a state of affairs in which promises are kept sees the state of affairs represented by 'p' as a goal that is to be realised in any (morally) perfect world.

Consider another sentence in natural language: 'it is wrong to tell lies'. What is the underlying form of this sentence in the new formal language? Blackburn no longer wants to make use of the B!-operator so it cannot be: 'B! (telling lies)'. An agent who disapproves of lying can be seen as approving of the obtaining of a certain state of affairs: namely a state of affairs in which lies are not told. This state of affairs is represented by the sentence: 'lies are not told'. The underlying form of 'it is wrong to tell lies' is therefore: 'H!p' where 'p' is the sentence 'lies are not told'.⁵⁴ An agent who disapproves of lying is committed to the view that a possible world is (morally) perfect just in case the proposition expressed by 'p' is true in that world.

'T!p' is, in Blackburn's new formal language, definitionally equivalent to '¬H!¬p'. 'T!p' expresses the view that the state of affairs represented by 'p' is consistent with, but not necessary for, a perfect world. Blackburn's general idea is that just as well as 'Op' can be seen as a wff in standard deontic logic, 'H!p' and 'T!p' can be seen as wffs in a logical system of expressions of attitude: "Thus, in the language to come, H!p is to be treated as a well-formed formula capable of entering the same embeddings as p" (ibid.:192).

⁵⁴ The underlying form can also be taken to be 'H!¬p' where 'p' is the sentence 'lies are told'.

3. Features of the Language

The formal logic that Blackburn develops for expressions of attitude is, to a large extent, modelled on a system of deontic logic developed by Hintikka.⁵⁵ Blackburn's development of Hintikka's system gets under way with the suggestion that we consider a set of sentences, L , that contains sentences with the 'H!'- and 'T!'-operator applied to them (let us call these sentences 'attitude-expressing' sentences) as well as declarative sentences to which no attitude-operators are applied (let us call these sentences 'state-describing' sentences). If, for example, L is $\{H!p, T!q, \neg p\}$, then the first two sentences of the set can intuitively be seen to represent the attitudinal commitments of a given agent whereas the last sentence can be seen to represent a state of affairs of the world in which the agent lives. L is therefore to be interpreted as a partial description of a possible world. With this interpretation of L in place, L tells us that the agent, whose attitudinal commitments are represented by 'H!p', and 'T!q', lives in a world that is less than ideal from his perspective. He takes the state of affairs represented by 'p' to be a goal, but he lives in a world in which ' $\neg p$ ' is realised.

Relative to L there is a morally ideal world: namely the world in which all the things that are taken to be goals are realised. To take a simple example: if L is $\{H!p, q\}$, then a world characterised by $\{H!p, p\}$ is a morally ideal world relative to L . The new world is a perfect world since everything that is taken to be an ideal in the original world is realised in the new world. Blackburn now introduces four important notions:

- 'a next approximation to the ideal, L^* , of L ' (ibid.:194). A next approximation, L^* , to the ideal of L is defined as a set of sentences that are brought about by applying the following rules to the sentences of L :

⁵⁵ See (Hintikka 1969).

- 1) If $H!A \in L$, then $H!A \in L^*$
- 2) If $H!A \in L$, then $A \in L^*$
- 3) If $T!A \in L$, then a set L^* containing A is to be added to the set of next approximations for L
- 4) If L^* is a next approximation to the ideal relative to some set of sentences L , then, if $A \in L^*$, then $A \in$ subsequent approximations to the ideal L^{**} , L^{***} ..

- 'a set of final ideals': "a set of final ideals, $\{L^{***} \dots\}$ of L is obtained when further use of these rules produces no new sentences not already in the members of L^{***} of the set" (ibid.:194).

- 'a route to an ideal': "..to each branch of a disjunction there corresponds a route to an ideal" (ibid.:194).

- 'unsatisfiability of a set of sentences': "A set of sentences L is unsatisfiable iff each route to a set of final ideals S results in a set of sentences S one of whose members contains both a formula and its negation" (ibid.:194).

Let me try to explain these four notions by relating them to a concrete example. Let L be $\{H!p \rightarrow H!q, T! \neg q, H!p\}$. These three sentences should be taken to constitute the root of a tree. The tableaux rules for the connectives of propositional logic apply to the connectives of Blackburn's logic. Importantly, this means that conditionals are treated as disjunctions. The first of the sentences generates branching, so we get a tree that branches. At the end of the left hand branch, we get $T! \neg p$ and at the end of the right hand branch, we get $H!q$. The set of sentences on the left hand branch is $\{T! \neg q, H!p, T! \neg p\}$ and the set of sentences on the right hand branch is $\{T! \neg q, H!p, H!q\}$. One first next approximation to L on the left hand branch is L^*1 . This is the set $\{H!p, p\}$. A second first next approximation to L on this branch is L^*2 . This is

the set $\{\neg q, H!p, p\}$. A third first next approximation to L on this branch is L^*3 . This is the set $\{\neg p, H!p, p\}$.

The first next approximation to L on the right hand branch of the tree is L^*4 . This is the set $\{H!p, p, H!q, q\}$. The second first next approximation to L on this branch is L^*5 . This is the set $\{\neg q, H!p, p, H!q, q\}$. Here is the complete tree:

	H!p→H!q, L	
	T!¬q, L	
	H!p, L	
T!¬p, L		H!q, L
H!p, L*1		H!p, L*4
p, L*1		p, L*4
		H!q, L*4
¬q, L*2		q, L*4
H!p, L*2		
p, L*2		¬q, L*5
		H!p, L*5
¬p, L*3		p, L*5
H!p, L*3		H!q, L*5
p, L*3		q, L*5
X		X

A set of sentences on the tree can intuitively be seen to represent a world. On the left hand branch, the set $\{L^*1, L^*2, L^*3\}$ is a set of final ideals relative to L . A set of final ideals is a set of worlds. Each set of sentences (i.e. each world) on this branch is a final ideal relative to L . Why is the set $\{L^*1, L^*2, L^*3\}$ a set of final ideals relative to L ?

The answer to this question is that if you apply the rules to the sentences of L^*1 , then you get a set of sentences, L^{**1} , that has the same sentences as L^*1 . If you apply the rules to the sentences of L^*2 , then you get a set of sentences, L^{**2} , that has the same sentences as L^*2 and if you apply the rules to the sentences of L^*3 , then you get a set of sentences, L^{**3} , that has the same set of sentences as L^*3 . Your new set $\{L^{**1}, L^{**2}, L^{**3}\}$ contains no sentences that do not already exist in the set $\{L^*1, L^*2, L^*3\}$. The latter set is therefore a set of final ideals relative to L . What is it for a set of sentences (a world) to be a final ideal relative to L ? Simply to be part of

a set of final ideals relative to L. On the right hand branch of the tree, $\{L^*4, L^*5\}$ is a set of final ideals relative to L and each set of sentences in the set is a final ideal relative to L.

Now, to the explanation of 'unsatisfiability'. Firstly, it should be noted that a route to a set of final ideals corresponds to a branch on a tree. Returning to our main example, L is unsatisfiable. The reason for this is that each route to a set of final ideals involves a set of sentences that contains a sentence and the negation of the sentence. On the left hand branch of the tree, this set is L^*3 . This set of sentences contains both $\neg p$ and p . On the right hand branch of the tree, this set is L^*5 . This set of sentences contains both $\neg q$ and q .

Blackburn's four rules are not completely self-explanatory, so a few comments on them seem needed. Rule 1) captures the idea that if something is considered to be a goal by an agent, then it continues to be a goal in the world that is a next approximation to the ideal relative to L. If one takes it to be a goal that, for example, promises are kept, and one lives in a world (L) in which it is not the case that promises are kept, then the goal continues to exist in a world (L^*) that is an approximation to the ideal of L in the sense that promises are kept in this world. Rule 2) specifies the intuition that if something is a goal, then this goal must be realised in any world that is a next approximation to the ideal of L.

Rule 3) has some similarity with the rule for the possibility-operator in standard systems of modal logic. If 'T!A' belongs to L, then rule 3) says that A is compatible with, but not necessary for, a perfect world. Consider the original tree. L is, on the left hand branch, $\{T!\neg q, H!p, T!\neg p\}$. Since $\neg q$ and $\neg p$ are not necessary for perfection, there is a morally perfect world relative to L in which neither $\neg q$ nor $\neg p$ is realised. On the tree, this world is L^*1 . Rule 3), however, says that there is at least

one development (world) in which what is tolerated in L is realised. In L there are two tolerations, so the left hand branch of the tree contains a world in which the first toleration is realised, L*2, and a world in which the second toleration is realised, L*3.

Blackburn's logical system is supposed to provide the means by which it is possible to check whether or not a set of sentences is consistent. The notion of consistency that Blackburn has in mind is, in Hale's interpretation, this:

"What matters for the consistency of L is that there should be at least one set of final ideals providing for the realisation, separately, of each of the T!A in, or implied by L, compatibly with realisation of all the H!B in or implied by L" (Hale 1993a:346).

Let me try to spell this out in a little more detail. Consider again L, on the left hand branch of the original tree. L contains two tolerations. The second toleration should be realised in a world different from the world in which the first toleration is realised. The intuition behind this is the following: I might take p to be a goal, tolerate that $\neg q$ and tolerate that $\neg p$. The fact that I tolerate that $\neg q$ and tolerate that $\neg p$ does, however, not mean that I tolerate $\neg q$ and $\neg p$ together. So, in order to check whether or not my attitudes are consistent I should not consider a world in which both of my tolerations are realised together. I should first consider a world in which my first toleration is realised and then see if that world is consistent when whatever it is that I take to be a goal is realised in that world. After that, I should consider a different world in which the second of my tolerations is realised and then see if that world is consistent when whatever it is that I take to be a goal is realised in that world. My attitudes are consistent only if both the two worlds are consistent. This procedure is

followed on the original tree by opening up the two worlds L^*2 and L^*3 . L is inconsistent since L^*3 is inconsistent.

The idea behind rule 4) can be brought into the open by taking L to be $\{H!H!p, H!q\}$. There is only one first next approximation to the ideal of L : namely L^* . L^* is $\{H!H!p, H!p, H!q, q\}$. L^* is not a final ideal relative to L . Application of the rules to the sentences of L^* , yields a further approximation to the ideal of L : namely L^{**} . This is the set $\{H!H!p, H!p, p, H!q, q\}$. L^{**} is a final ideal relative to L , i.e. a morally perfect world relative to L . If rule 4) was not in place, then 'q' would not carry into L^{**} from L^* . A consequence of this would be that a morally perfect world relative to L , would not contain the realisation of one of the ideals that obtain in L , and such a world could hardly be said to be a morally perfect world relative to L .

If these considerations help to make clear the rationale behind 4), then Blackburn has a comment on 4) that, at least to my mind, somewhat perplexes the picture: "If L^* is already a next approximation to the ideal and contains a sentence A , then except where A derives from realization of a toleration, it must transfer to further approximations to the ideal L^{**} .." (Blackburn 1988:195).

Let L be $\{H!H!p', T!q\}$. One first next approximation to the ideal of L is L^*1 . This is the set $\{H!H!p, H!p\}$. A second first next approximation to the ideal of L is L^*2 . This is the set $\{H!H!p, H!p, q\}$. Application of the rules to the sentences of L^*1 gives L^{**1} which is the set $\{H!H!p, H!p, p\}$. Application of the rules to the sentences of L^*2 gives L^{**2} which is the set $\{H!H!p, H!p, p, q\}$. $\{L^{**1}, L^{**2}\}$ is a set of final ideals relative to L . Rule 4) clearly says that 'q' should carry over from L^*2 into L^{**2} , but from what Blackburn says in the above passage, 'q' should not carry over into L^{**2} since 'q' in L^*2 derives from realisation of a toleration. There is,

in other words, an inconsistency between what Blackburn's rule 4) actually says and what Blackburn says as a comment to the rule.

In connection with these considerations about rule 4), a feature of rule 3) springs to mind. If rule 3) is to be understood literally, then a formula of the form 'T!A' in L^* should carry over into subsequent approximations. Rule 3) only says that if 'T!A' belongs to L , an initial set, then the toleration must be realised in at least one of the next approximations to the ideal of L . The rule is silent on what is to happen if the formula occurs at L^* or any next approximations to the ideal of L . The only rule one has to go by in this scenario is 4). Application of this rule, however, means that 'T!A' automatically carries over and is never realised. But if the toleration is never realised, then we will never get an answer to the question of whether or not an original set of sentences is consistent in the desired sense of 'consistent'. If a toleration is implied by one of the formulae of L , but does not occur explicitly in L , as it is the case in, for example, 'H!($p \rightarrow T!q$)', then this toleration must be realised at some point in order to see whether L is consistent.

Blackburn seems to be in agreement with this. As a comment to rule 3), he says: "The rule is that if T!A is present in a set, then there must be a next set in which A is present, although it is not to be in all" (ibid.:194). Blackburn here talks generally about a set of sentences, and not only about the set of sentences that is the initial set.

My worry about the formulation of 3) also has force against the formulation of rule 2). 2) only says that if H!A belongs to the initial set, then A is realised in any first next approximation to the ideal of L . However, what Blackburn means is that if H!A is present in a set, not only the initial set, then A must be present in any set that is a further approximation to the ideal of L . At one place, Blackburn says: "We want to iterate the procedure of generating a next ideal. This can be done by repeated use

of these rules" (ibid.:195). The idea is that if the set $L^* \{H!H!p, H!p, H!T!q, T!q\}$ is the result of applying the rules to the initial set, then the rules are to be reapplied to this set in order to generate a further approximation to the ideal of L. My interpretation of Blackburn is that both rule 2) and 3) are to be applied to this set. Blackburn certainly thinks that rule 2) should be applied, and since there is nothing in the formulation of the rules that suggests that rule 3) is different from rule 2) in this respect, I assume that rule 3) should also be applied. If rule 2) and 3) are different in this respect, it is a mistake not to state that clearly in the formulation of the rules.

4. *Validity, Consistency and the Merits of the Language*

After this account of the four rules that are to be applied in order to generate next ideals, attention can now be directed at the notion of validity. I here quote all that Blackburn says on this important issue:

"Logical truth of A is unsatisfiability of the negation of A; B is a logical consequence of A if and only if $A \rightarrow B$ is valid, that is $(A \& \neg B)$ is unsatisfiable" (ibid.:193).

If I understand Blackburn correctly, then the definition of validity is a definition of semantic validity. To repeat a point made earlier: presented with the question of what the distinctive feature is of a semantically valid inference, Blackburn cannot say that what is special about it is that it is impossible that its premises should be true and its conclusion false. In other words, he cannot say that a semantically valid inference is necessarily truth-preserving. The quoted passage suggests that Blackburn would say that one tests an inference in natural language for semantic validity by checking whether or not a certain set of sentences in the formal language is unsatisfiable. What

sentences should go into the this set? The premises of the inference and the negated conclusion. The inference is semantically valid just in case this set is unsatisfiable.

How might Blackburn's formal language and definition of validity help him deal with Geach's original modus ponens inference? In AC, Blackburn no longer holds the view that conditionals with evaluative constituent sentences serve the semantic function of expressing second order attitudes. The underlying form of 'If it is wrong to tell lies, then it is wrong to get your little brother to tell lies' is not interpreted as being 'H!(/B!(telling lies)/;/B!(getting your little brother to tell lies)/)' but something much simpler, namely: 'H!p→H!q'. The underlying form of Geach's inference is therefore now seen to be this: 'H!p→H!q, H!p ⊢ H!q'. Is the MPP inference in natural language valid? The set of sentences, L, that is to be checked for satisfiability is {H!p→H!q, H!p, T!¬q}. This is the same set of sentences as in the originally worked out example in section three. We already know that this set is unsatisfiable. It can therefore be concluded that the MPP inference is valid.

One of the advantages of Blackburn's logical system is that it enables him to deal with patterns of inference other than modus ponens. If the expressivist has a problem with Geach's inference, then he also has a problem with an inference such as this: 'either it is right to keep promises or it is right to tell the truth; it is not right to keep promises; therefore, it is right to tell the truth'. There is a very strong appearance that this is a valid inference in natural language. Blackburn's proposal in *Spreading the Word* as to how the expressivist should deal with Geach's inference is a proposal with very narrow scope in the sense that it cannot, or at least cannot obviously, be used to deal with other patterns of inference. It is difficult to see how an agent affirming the major premise of the above inference could properly be seen as expressing a second order attitude towards a certain combination of first order

attitudes. So, even if Blackburn's proposal in his earlier work was entirely successful, he would still be faced with the task of making sense of patterns of inference different from modus ponens. Translated into the formal language, the above inference will be of the form: $H!p \vee H!q, T!\neg p \vdash H!q!$. The set of sentences, L , that is to be checked for satisfiability is $\{H!p \vee H!q, T!\neg p, T!\neg q\}$, and it can be shown that this set is unsatisfiable.

Let me try to make clear the link between the notion of 'unsatisfiability of a set of sentences' and the notion of 'inconsistency of a set of attitudes'. Consider an agent who accepts the premises but not the conclusion of Geach's inference. He has a set of attitudinal commitments that is represented by this set of sentences: $\{H!p \rightarrow H!q, H!p, T!\neg q\}$. Depending upon how he resolves the conditional, the agent will have either of the following sets of attitudes: $\{H!p, T!\neg q, T!\neg p\}$, $\{H!p, T!\neg q, H!q\}$. Both of these sets are inconsistent. If the agent has the first set, then he takes $\neg p$ to be compatible with a morally perfect world but another of his attitudes is that p must be realised in a world if that world is to be perfect. But then the agent is committed to the view that there is at least one possible world that is morally perfect and in which $\neg p$ and p are both realised. But such a world is logically impossible. The same type of problem arises in relation to the second set.

Consider now the fact that the set of sentences $\{H!p \rightarrow H!q, H!p, T!\neg q\}$ is unsatisfiable as is evident from the closed tree of section three. The fact that this set is unsatisfiable is evidence that the set of attitudes that this set of sentences can be taken to represent, is inconsistent. Why is this? We know from the definition of consistency of a set of attitudes that such a set is consistent just in case it meets both of the following conditions: 1) for each of the tolerations in the set there is a logically

possible world in which that toleration is realised and in which everything endorsed in the set is realised: 2) there is a logically possible world in which everything endorsed in L is realised. If such a set of consistent worlds exist, then the set of attitudes represented by the sentences of L is consistent.

The tree of section three can be interpreted as an attempt to check whether or not a set of attitudes represented by the sentences of L meets conditions 1) and 2). The set of attitudes represented by the L-sentences on the left hand branch of the tree is inconsistent because what happens if we try to realise in a world, L*3, one of the tolerations of L and all the endorsements of L, is that we get a logically impossible world. The syntactical fact that the set of L-sentences on this branch is unsatisfiable can be taken to show that the set of attitudes represented by these sentences is inconsistent. Similarly, on the right hand branch of the tree. The L-sentences here constitute a set of sentences that is unsatisfiable. This is a syntactical fact. It can, however, be taken to show that the set of attitudes represented by L is inconsistent. L*5 is namely a world in which we try to realise a toleration of L and all the endorsements of L, and what we get is a logically impossible world.

Consistency in goals or attitudes is a cardinal virtue for Blackburn as is evident from these passages:

"There is nothing surprising about using realisations of goals or ideals as the final test for consistency. The ordinary way of finding whether recommendations are consistent is to imagine them carried out and see if this can be consistently done" (ibid.:189).

"In turn this gives us a needed notion; a person may be something worse than 'immoral', or possessing contingently defective attitudes, but not be

inconsistent in the sense of believing anything logically false. He may simply have ideals that admit of no consistent realisation" (ibid.:196).

The content of a more informal answer to the question of what the distinctive feature is of a valid pattern of inference, can now be seen. On Blackburn's account, Geach's inference is valid because acceptance of its premises and failure to accept its conclusion saddles an agent with an inconsistent set of attitudes. More generally, Blackburn will say that when it comes to valid inferences whose constituent sentences are evaluative (non-truth-apt), the designated semantic value that is preserved from premises to conclusion is consistency in attitude.

5. A New Semantics For Connectives

If all the moves that Blackburn has made so far are granted him, then he is in a position to show that the set of sentences that represents the attitudinal commitments of an agent who accepts the premises but not the conclusion of evaluative modus ponens, modus tollens and disjunctive syllogism inferences, is an unsatisfiable set. In other words, Blackburn is in a position to show that an agent ends up with inconsistent attitudes if he accepts the premises but not the conclusion of inferences that are instantiations of the above patterns of inference.

Before being too confident on Blackburn's behalf, there are, however, various points about his account that needs to be addressed. The most important question is perhaps the one of how the connectives of the formal language are to be understood. One thing that is immediately clear is that Blackburn cannot treat the connectives of his formal language in the usual truth-functional way. Formulae like 'H!A' and 'T!B' are, in Blackburn's formal language, given an expressive interpretation. Their function is not to express propositions but to express attitudes, and as a result of that

they do not have a genuine truth-value. Given that 'H!A' and 'T!B' function expressively, how are we to understand complex expressions in which formulae of the above kind lie within the scope of the classic sentence-operators?

Blackburn is aware that he needs to address this question. Let us begin with Blackburn's account of the cases in which formulae that function expressively lie within the scope of the negation sign. As noted in section two, 'H!p' expresses approval of its being the case that p. '¬H!p' expresses toleration of ¬p. The intuitive explanation of this is the following: an agent who denies that a world is morally perfect just in case p is realised in that world, sees the realisation of ¬p as being consistent with a world being morally perfect. Such an agent tolerates, so to speak, the realisation of ¬p. Now, toleration of something can, as we know by now, be expressed with the use of the T!-operator. So, 'T!¬p' is substitutionable for '¬H!p'. According to Blackburn, the two formulae express, if not exactly the same attitude, then at least an almost identical attitude.

In the former formula, the scope of the negation is a sentence that expresses a proposition, whereas the scope of the negation of the latter formula is a sentence that do not express a proposition. This feature of the latter formula might, Blackburn acknowledges, make us uneasy but the fact that it, and every other instance of external negation, can be converted into a formula where the scope of negation is a proposition-expressing sentence, should, according to Blackburn, dissipate any such unease (ibid.:192).

When Blackburn is to give an interpretation of formulae in which sub-formulae that are expressive of attitude occur within the scope of binary connectives, Blackburn opts for a different strategy than the one he uses to give an interpretation of formulae in which sub-formulae that are expressive of attitude lie

within the scope of negation. The general strategy for dealing with formulae of the former kind draws upon the way formulae with a dominant binary connective are standardly treated in a sentence tableaux of propositional logic. Ordinary evaluative talk contains sentences such as 'Jones is innocent or Jones should be punished'. The underlying form of this is ' $p \vee H!q$ ' where ' p ' stands for 'Jones is innocent' and ' q ' stands for 'Jones is punished'.

Blackburn offers an expressive interpretation of the disjunction. An agent who advances it, does not say anything that can be evaluated in terms of truth, but expresses a special kind of commitment. Blackburn brands this 'a tree-tying account' of a disjunction with an evaluative disjunct. The idea is that by advancing the disjunction, an agent ties himself to a tree of possibilities. The analogy is that in propositional logic, an agent who affirms ' $p \vee q$ ' is tied to a tree of possibilities. It is not a possibility for such an agent, if he is to avoid being inconsistent, to go on to affirm ' $\neg p$ ' and ' $\neg q$ '. If the agent goes on to do just this, then he is inconsistent in the sense that he holds beliefs that cannot all be true.

One can be in a certain non-cognitive state to the effect that one is prepared to accept either of two commitments, where the notion of a 'commitment' is taken to be broad enough to cover both a regular belief and an attitudinal stance, and it is this psychological state that one expresses by advancing the disjunction. Blackburn's tree-tying account of evaluative disjunction amounts to the development of what can be called a 'commitment-semantics' for disjunction. Blackburn's idea is that the meaning of this sentence-operator can be explained by spelling out what kind of commitments an agent incurs if he asserts a sentence in which this sentence-operator is the main connective. Basically, one who asserts the above disjunctive sentence about Jones is committed to do either of two things: accepting the proposition

expressed by the first disjunct or endorsing the attitude expressed by the second disjunct.

Blackburn repeats his tree-tying account when he is to explain the meaning of the 'if.., then..' sentence-operator in cases where this connective binds together sentences that do not both have a (genuine) truth-value. ' $A \rightarrow B$ ' is, in standard propositional logic, equivalent to ' $\neg A \vee B$ ' and Blackburn takes it to be the case that the commitment incurred by asserting an evaluative conditional is a commitment to a disjunction. If one, for example, asserts the conditional 'if organised sports games are good, then organised sports games ought to be part of the school curriculum' ($H!p \rightarrow H!q$), then one is committed to either of two attitudes: tolerating that it is not the case that organised sports games take place or approving of it being the case that organised sports games are part of the school curriculum. So by asserting the conditional, an agent is, according to Blackburn, tied to a tree of possibilities concerning what attitudes to adopt. The fact that he is so tree-tied can, furthermore, be used to assess the consistency of his attitudes. Suppose that the agent goes on to accept that organised sports games are good (accepts ' $H!p$ ') but does not accept that organised sports games ought to be part of the school curriculum (does not accept ' $H!q$ '). Then he has accepted the premises of a moral modus ponens but failed to accept its conclusion and we know already that no matter how the agent resolves the conditional, he ends up with an inconsistent set of attitudes.

6. A Problem With Rule 4)

I hope that the previous sections constitute an acceptable account of the main ideas of AC. The question that now needs to be answered is the question of what we should think of Blackburn's new ideas. One cause for concern is Blackburn's rule 4). In my

account of this rule, I said that there is an inconsistency between what rule 4) actually says and what Blackburn says as a comment to the rule. This inconsistency is well explained by the fact that Blackburn, as noted by Hale, faces a dilemma concerning rule 4).⁵⁶ The dilemma is this: with rule 4) in place ' $H!T!p \vdash H!p$ ' is valid since $\{H!T!p, T!¬p\}$ is unsatisfiable. But the fact that this inference is valid is an unwanted result. This means that if you endorse tolerating that p , then you are committed to endorsing that p . But one can surely endorse the toleration of, say, the occurrence of a rally in favour of the local communist party without endorsing the occurrence of the rally.

However, an amendment of the rule along the lines suggested by Blackburn, has as a result that ' $\vdash H!(H!p \rightarrow p)$ ' is not valid.⁵⁷ What you get when you negate this sentence is the sentence ' $T!(H!p \ \& \ ¬p)$ ', and this sentence can only be shown to be unsatisfiable with the original, unrestricted, rule 4) in place.

To this, one might reply: 'What is so bad about the fact that ' $\vdash H!(H!p \rightarrow p)$ ' is not valid? Isn't the solution to the dilemma just to adopt the amended rule and accept that the above inference is not valid?' I do not think so. If ' $\vdash H!(H!p \rightarrow p)$ ' is not valid, then ' $T!(H!p \ \& \ ¬p)$ ' is satisfiable. This means that in Blackburn's logical system for expressions of attitude, it is consistent to have the attitude ' $T!(H!p \ \& \ ¬p)$ '. But a logical system that allows that attitude to be consistent cannot be right. If anything is intolerable from the perspective of a given agent, it is surely a situation in which that agent approves of something, and the negation of what he approves of, is the case. An agent with the attitude expressed by ' $T!(H!p \ \& \ ¬p)$ ', however, tolerates such a situation. Such an agent has no proper grasp of the concept of toleration. His

⁵⁶(Hale 1993a:348).

⁵⁷The amendment is suggested in (Blackburn 1988:195), and the passage where it occurs is quoted in section three.

attitudinal stance is not so much strange from a first order moral perspective: rather, it is conceptually incoherent. We want this to be reflected in a proper logic of attitudes and we therefore want $\vdash H!(H!p \rightarrow p)$ to be a theorem in such a logic. So, I think the dilemma facing Blackburn in relation to rule 4) is real enough.

The existence of the dilemma shows that rule 4) is neither acceptable in its original form nor in its amended form. The weakest conclusion to be drawn from this is that further work needs to be done on rule 4). As a response to Hale's criticism, Blackburn suggests that 4) should be substituted with:

4') If L^* is a next approximation relative to some set of sentences L , then if L^* contains $H!(X)$ then a subsequent approximation L^{**} contains X and all the other sentences of L^* (Blackburn 1993b:381).

With this rule in place, $\vdash T!(H!p \ \& \ \neg p)$ can be shown to be unsatisfiable. L^* is $\{H!p, \neg p\}$. Since L^* contains $H!p$, which is an instantiation of the form $H!(X)$, L^{**} contains, in addition to $H!p$ and $\neg p$, p . $\{L^*, L^{**}\}$ is a set of final ideals relative to L . This set contains a set of sentences that involves a sentence and its negation. L is therefore unsatisfiable and $\vdash H!(H!p \rightarrow p)$ is therefore valid.

What about the first example? Does 4') secure that $H!T!p \vdash H!p$ is invalid? Blackburn thinks so: "Certainly this blocks the unwanted result that Hale cites. Everything stops where it should, at $\{T!p, \neg p\}$ " (Ibid.:381). The set of sentences that is to be checked for satisfiability is $\{H!T!p, T!\neg p\}$. L^{*1} is $\{\neg p, H!T!p, T!p\}$. There is, however, a further next approximation to the ideal of L . This is the set in which the toleration of L^{*1} is realised. L^{**1} is $\{p, H!T!p, T!p, \neg p\}$. Importantly, $\neg p$ of L^{*1} carries over to L^{**1} because 4') says that if L^* contains something of the form $H!(X)$, then all the sentences of L^* carries over to subsequent approximations. In the

present example, $\neg p$ is a sentence of L^*1 , and this set contains $H!T!p$. L^{**1} therefore contains $\neg p$. This means that L^{**1} contains a sentence and its negation and $\{H!T!p, T!\neg p\}$ is therefore unsatisfiable. This has as a result that, contrary to what Blackburn thinks, when 4') is in place, $H!T!p \vdash H!p$ is still valid. The substitution of 4) with 4') should therefore not be considered to be a proper solution to the initial dilemma. With 4') in place, we still get an unsatisfactory result.

What could Blackburn reply to this? He could say that the ability to show that that $\{H!T!p, T!\neg p\}$ is unsatisfiable rests on a strong, and mistaken, reading of rule 3). In particular, it is because rule 3) is applied to one of the sentences of L^*1 that we get a new set of sentences, L^{**1} , that contains $\neg p$ and therefore get closure of the three. But perhaps this is a mistaken application of rule 3). A weak reading of 3) is that this rule is only to be applied to an initial set of sentences, and not to any sets of sentences that are next approximations to the ideal of L . On the weak reading of 3), $\{H!T!p, T!\neg p\}$ is satisfiable since $T!p$ of L^*1 never gets realised and we therefore never get a set of sentences that contains both $\neg p$ and p . As Blackburn says in the above quote, everything stops where it should: namely at $\{T!p, \neg p\}$.

What should the response be to Blackburn's suggestion that a weak reading of rule 3) should be adopted? One might argue against the suggestion on the ground that its adoption has as a result that an inference that intuitively seems to be valid cannot be proved with the available rules. Consider now the inference: $H!T!p \vdash T!p$. It seems to me that this inference is intuitively valid. Approval of tolerance of p seems to entail tolerance of p . I would be inconsistent in my attitudes if I, on the one hand, approved of toleration of, say, the occurrence of a rally in favour of the local

communist party, and then, on the other hand, did not tolerate the occurrence of the rally.

The inference cannot, however, be shown to be valid when 3) is given a weak reading. It can only be shown to be valid when 3) is given a strong reading. So, it is not, I conclude, a viable option for Blackburn to insist on a weak reading of 3). This rule must be given a strong reading. But then Blackburn's substitution of 4) with 4') does nothing to overcome Hale's dilemma. This dilemma still stands.

7. Further Critiques of the Formal Language

Let me now try to present an objection to Blackburn that turns on a feature of his formal language. As is evident from the following passage, Blackburn considers formulae involving reiterated H!-operators to be wffs in his formal language:

"to take a trivial case $H!H!H!p$ gives as an ideal the set $\{H!H!H!p, H!H!p, H!p, p\}$ (all the states commended are imagined realised)" (Blackburn 1993b:380).

If ' $H!H!p$ ' is a wff in Blackburn's formal language, then the sentence in natural language that this formulae is supposed to be a representation of, must be considered to be a meaningful sentence. Blackburn's formal language is supposed to give a precise mapping of natural language so that the deductive relationships between sentences in natural language can be studied in the formal language. But giving inference rules for the H!-operator, not only when it is applied to a sentence-letter but also when it is reiterated as in ' $H!H!p$ ', has no relevance for the project of casting light on the deductive relationships of sentences in natural language if the sentences that formulae involving reiterated H!-operators are supposed to be representations of, are considered to be meaningless. What is the point of making clear what can be

inferred from 'H!H!p' if the natural language sentence that this formulae is supposed to be a representation of, is meaningless? Blackburn's view that 'H!H!p' is a wff in his formal language therefore commits him to the view that the natural language sentence that this formulae is a representation of, is a meaningful sentence.

If this is accepted, then we can ask the question of what natural language sentence it is that 'H!H!p' is a representation of. 'H!p' is the formularisation of 'it is good that p'. It would therefore seem natural to think that 'H!H!p' is the formularisation of:

(1) It is good that it is good that p.

Blackburn is therefore, it seems to me, committed to the view that (1) is meaningful. The central question now is what (1) means on the expressivist picture. A good way of approaching this question is to look at what an expressivist could say about the meaning of the semantically simpler sentence:

(2) It is good that p.

One possible explanation of the meaning of (2) begins with the observation that a declarative sentence such as 'football is played in the park', can be uttered with various forces. To keep things simple, we can here just focus on assertoric, interrogative and expressive force. If 'p' is uttered with assertoric force, then the proposition expressed by 'p' is asserted. If 'p' is uttered with interrogative force, then no proposition is asserted: rather a question is being asked - namely the question of whether football is played in the park. If 'p' is uttered with expressive force, then no proposition is asserted: rather an attitude is expressed towards the state of affairs in which the proposition expressed by 'p' is true.⁵⁸

⁵⁸A distinction should here be made between two different senses of 'expresses'. It is sometimes said that a sentence expresses an attitude: "A parallel to the idea that a certain sentence expresses an attitude ('Hurray to the Bears')" (Blackburn 1988:186). Strictly speaking, what a sentence can express is a proposition. A sentence cannot express an attitude. People, not sentences, have attitudes that they

In spoken language, intonation can be a guide to the force of an utterance. This is especially clear in connection with interrogative force. 'p' uttered with a special intonation often indicates that a question is being asked, rather than something is being asserted. In written (natural) language, there exist signs that indicate what kind of force a sentence is being put forward with. As an example of this, consider the fact that '?' after a declarative sentence, as in 'p?', indicates that 'p' is put forward with interrogative force. A sincere agent who writes 'p?' has not thereby asserted a proposition: rather he has asked a question. What about assertoric force? There is no sign in conventional, written language that indicates that a sentence is being put forward with assertoric force. However, as has been noted, the fact that a declarative sentence in a non-fictional text stands by itself between full stops, indirectly indicates to the reader that a proposition is asserted.⁵⁹ Borrowing Frege's sign for assertion, '⊢', it is possible to make explicit that a sentence is put forward with assertoric force. An agent who writes '⊢ p' indicates thereby that he asserts the proposition expressed by 'p'.

Spoken natural language does not allow us, by a special intonation in the utterance of 'p', to indicate that the 'p' is uttered with assertoric force. The default position, however, is that if a declarative sentence is uttered by itself, i.e. in between full stops, and if it is not embedded in a larger sentence, as in, say, a conjunction or a conditional, then it is put forward with assertoric force. Regarding expressive force, there is no sign in written, natural language that indicates that 'p' is put forward with expressive force. In Blackburn's formal language, such a sign, however, exists. A sincere agent who writes 'H!p' in this language does not thereby assert a proposition

can express, and the sense of 'express' used in relation to people is different from the sense of 'express' used in relation to sentences expressing a proposition.

⁵⁹ See (Geach 1965:456).

(especially not the proposition expressed by the sentence 'I approve of p'). The agent rather expresses approval of the state of affairs in which the proposition expressed by 'p' is true. Writing 'H!p' is equivalent to writing something like 'Hurray to p'.

If Frege's assertion sign and Blackburn's sign for expressive force were adopted into conventional, written language, we could have these three sentences: '⊢p', 'p?' and 'H!p'. Each of these three sentences has its own distinctive meaning. Just to focus on the two latter sentences: to understand what an author was doing who wrote 'p?', it would be insufficient to know the truth-conditions of 'p'. To understand the meaning of 'p?', one would have to understand that a question is being asked. Similarly with 'H!p'. To understand what an author was doing who wrote 'H!p', it would be insufficient to know the truth-conditions of 'p'. To understand the meaning of 'H!p', one would have to know that by writing 'H!p', an author expresses approval of the state of affairs in which the proposition expressed by 'p' is true. The three sentences mirror some of the various things that we can do with a proposition: namely assert it, question its truth and approve of it being true.

Now, if conventional natural language does not allow us to indicate expressive force by use of a special sign, how do we then indicate this special force? The expressivist could here say that we do this by using the string of words 'it is good that'. As noted, 'H!p' is the proper formalisation of 'it is good that p'. In the formal language, 'H!' is an indicator of expressive force, so it is very natural to suggest that 'it is good that' indicates in natural language, what 'H!' indicates in the formal language. There is not anything essentially misguided about the idea of indicating force through use of a certain string of words. The interrogative force of an utterance can be indicated by placing the string of words 'is it the case that' before a declarative sentence. A sincere agent who writes 'p?' could just as well have written 'is it the case

that p '. Similarly, an agent who utters ' p ' with a special questioning intonation could just as well have uttered the words 'is it the case that p '.

Let us now return to the meaning of (2). On the assumption that the string of words 'it is good that' functions as a force indicator of expressive force, the expressivist could explain the meaning of (2) by saying that by uttering (2), a sincere agent expresses approval of the state of affairs in which the proposition expressed by ' p ' is true.

The view that the string of words 'it is good that' is a force indicator, however, commits Blackburn to the view that (1) is a sentence involving reiterated force indicators. This is not an attractive view to be committed to. In general, it is difficult to see what the meaning is of sentences involving reiterated force indicators. What does ' $\vdash \vdash p$ ' or ' $p??$ ', for example, mean? Sentences such as these do not have a conventional meaning. There is no conventional use of these sentences in natural language that establishes the meaning of the sentences. It is, of course, open to a theorist to put forward a stipulative definition that makes clear what, say, ' $p??$ ' means in his use of the sentence and makes clear how he thinks that ' $p??$ ' should be used. So, it is not the case that sentences involving reiterated force indicators are intrinsically meaningless.

This might provide an opening for Blackburn. Perhaps Blackburn can just put forward a stipulative definition of what he takes (1) to mean? In my opinion, however, this is not a proper response to the original problem. Blackburn's formal language in which ' $H!H!p$ ' is a wff is not just any language. As already noted, it is a formal language that is supposed to provide a formal mapping of natural language as it is actually spoken. Giving a stipulative definition of the meaning of (1) as a response to the question of what (1) means, is, in effect, just to admit that (1) does

not have a conventional meaning in natural language and if (1) does not have a conventional meaning in natural language, then there is no case for thinking, as Blackburn does, that the formula in the formal language, that is the representation of (1), is a wff.

I think that these considerations show that Blackburn is ill advised to explain the meaning of (1) by construing the string of words 'it is good that' as a force indicator of expressive force. Taking into account the fact that the H!-operator in the formal language functions as a force indicator of expressive force, it, however, seems that this is the explanation that he is committed to.

Let us, for the sake of the argument, leave this problem to one side and allow the expressivist to give an alternative explanation of the meaning of (1): an explanation that, importantly, does not have to cohere with the role played by the H!-operator in the formal language. What might such an alternative explanation look like?

Let us again begin with (2). The expressivist could here offer the following analysis: 'it is good that' is a piece of 'expressive' vocabulary that can be put in front of a proposition-expressing sentence to form a new sentence such that when a sincere agent utters the new sentence, 'it is good that p', he expresses an attitude of approval towards the state of affairs described by the proposition-expressing sentence. Such an analysis coheres well with what Blackburn says in this passage:

"Normally, if I make plain to you what I feel, say about the Bears, I will most probably do so using a sentence with an 'expressive' predicate: 'the Bears are great!'" (Blackburn 1988:187).

At this level of semantic complexity, the expressivist analysis is intelligible: but what happens if we move a step up on the ladder of semantic complexity and arrive a (1)?

One possible way for the expressivist to analyse (1) consists in repeating the analysis given of (2). This would involve saying that 'it is good that p' is a sentence that expresses a proposition and that 'it is good that' can be put in front of this sentence to generate a new sentence such that when an agent utters this new sentence, he expresses approval of the state of affairs described by the proposition-expressing sentence 'it is good that p'. An expressivist cannot, however, give this analysis since it involves accepting that 'it is good that p' is a sentence that expresses a proposition and this is something that can only be accepted at the price of abandoning expressivism.

This shows that expressivism is open to the critique that it offers no uniform account of the semantic role of 'it is good that'. In (1), this string of words does not attach to a sentence expressing a proposition to generate a new sentence such that when an agent utters this new sentence, he expresses approval of the state of affairs described by the proposition-expressing sentence. However, this is exactly what the string of words does in (2). The meaning of (1) is therefore still left unexplained.

A more radical way in which Blackburn could reply to the objection that he cannot render (1) meaningful, consists in denying that (1) is meaningful. This is actually what I think he should do. (1) is a grammatically correct sentence, but it is not obvious (at least not to me) that it has a clear meaning. The problem is, however, that if Blackburn denies that (1) is meaningful, then 'H!H!p' is not a wff in his formal language. Two options are now open to Blackburn. The first consists in biting the bullet and acknowledging that formulae involving reiterated H!-operators are not wffs in his formal language. The second consists in rejecting my suggestion that (1) is the natural language sentence that 'H!H!p' is the representation of. It is some other

natural language sentence that 'H!H!p' is the representation of and this other sentence is, contrary to (1), meaningful.

If Blackburn takes the first line, then it is, however, impossible to see how it can be the case, let alone a trivial case, as Blackburn insists it is, that H!H!H!p gives as an ideal the set {H!H!H!p, H!H!p, H!p, p}. This case involves a formulae that is not a wff, and it is surely absurd to exemplify the inference rules of a logical system by making use of a sentence that is not a wff in the formal language of that system.

If Blackburn takes the second line, then it becomes very difficult to understand what the H!-operator means in Blackburn's formal language. As previously noted, the sentence in natural language that 'H!p' is a representation of, is 'it is good that p'. How this could be consistent with the denial that H!H!p is a formularisation of (1), I simply fail to see. In general, I therefore think that there is no obvious way for Blackburn to escape the problem that his admission of the well-formedness of 'H!H!p' presents him with.

8. *Concluding Remarks*

I think that the objections pressed in the previous two sections show that Blackburn's logical system is unsatisfactory as it stands. However, if we leave to one side the technical, or semi-technical, problems discussed in these sections, then it seems to me that Blackburn, with the introduction of the idea of a commitment-semantics for binary connectives, has made real progress, at least in comparison to the ideas of *Spreading the Word*, in relation to the two important questions of what, on the expressivist picture, sentences like Geach's conditional mean, and what the validity of, say, a moral modus ponens consists in. This idea seems promising, and it should be pursued and developed further.

Chapter 7

Round Three of the Frege-Geach Point

1. *Introductory*

In a number of recent writings, Blackburn has refined his commitment-theoretic account of evaluative compounds.⁶⁰ He has also argued that what validates an evaluative instantiation of a standardly accepted pattern of inference, is the fact that accepting its premises and failing to accept its conclusion, requires incurring an inconsistent bundle of commitments.⁶¹ It is a central part of Blackburn's commitment-semantics to give the commitment condition for each of the sentence-operators in natural language. To give the commitment condition for, say, conditionals consists in spelling out precisely what complex commitment it is that a speaker incurs by accepting an evaluative conditional. Bob Hale has offered a critique of Blackburn's commitment-theoretic account of evaluative conditionals (Hale 2002).

In sections two to five of this chapter, I present this critique and give an account of Blackburn's reply to it. In section six, I argue that Blackburn's reply is flawed because the account of conditionals and disjunctions that is involved in this reply, fails to validate all standardly accepted patterns of inference. In the last section of the chapter, I place the relatively technical point of section six in a broader perspective: I argue that this point seriously diminishes the plausibility of

⁶⁰See (Blackburn 1998; 2002).

⁶¹"Logic is our way of codifying and keeping track of intelligible combinations of commitments" (Blackburn 1998:72). Blackburn's definition of (semantic) validity is this: an inference is valid just in case acceptance of its premises together with failure to accept its conclusion result in an inconsistent set of commitments.

Blackburn's overall quasi-realist project. The chapter ends with some concluding remarks on expressivism and the Frege-Geach point.

2. First Proposal

Two sorts of conditionals fall under the concept of an evaluative conditional: a) conditionals in which the antecedent and the consequent are attitude-expressing, and b) conditionals that have either an attitude-expressing antecedent or an attitude-expressing consequent ('or' here represents an exclusive disjunction). It is impossible to formulate the commitment condition for evaluative conditionals in full generality if a term cannot be found that is neutral between the two types of commitment that are incurred by direct affirmation of the constituent sentences of a conditional of type b). The term 'accept' may be employed here (Hale 2002:146). In full generality, the commitment condition for evaluative conditionals is therefore this: to accept ' $p \rightarrow q$ ' is to incur a commitment to:

- (1) either not accept that p or accept that q.

Where 'p' expresses a proposition, 'a commitment to accept that p' means 'a commitment to believe the proposition that is expressed by direct affirmation of p'.

Where 'q' is an attitude-expressing sentence, 'a commitment to accept that q' means a 'commitment to endorse the attitude that is expressed by direct affirmation of q'.

The point now is that this commitment condition for a conditional enables an expressivist to show that a modus ponens inference involving an evaluative conditional is valid. Consider an agent who accepts the premises of such an inference but fails to accept its conclusion. Such an agent must have either of the following two sets of commitments: i) accepting that p, not accepting that p and not accepting that

q: ii) accepting that p, accepting that q and not accepting that q.⁶² Either way the conditional is resolved, the agent ends up with inconsistent commitments. He either accepts that p and does not accept that p or he accepts that q and does not accept that q.

Given the earlier definition of validity, and given the fact that an agent who accepts the premises but not the conclusion of a modus ponens inference, ends up with inconsistent commitments, it can be concluded that this pattern of inference is valid.

As Hale notes, an expressivist who adopts a commitment-semantics must hope that the semantics makes valid, not only a modus ponens inference, but all the standardly accepted patterns of inference (ibid.:147). In line with Hale's way of proceeding, let us take a modus tollens inference as an example. Does the commitment-semantics make it the case that an agent is inconsistent if he accepts the premises but not the conclusion of such an inference?⁶³

Let the conditional of the inference be an evaluative conditional of type b). The conditional can be resolved in two different ways, so depending on how the conditional is resolved, the agent is either committed to: i) accepting that $\neg q$, accepting that q and not accepting that $\neg p$: or ii) accepting that $\neg q$, not accepting that p and not accepting that $\neg p$. In the latter case, the agent does not accept that p and he does not accept that $\neg p$. i) is clearly an inconsistent set of commitments. It involves both accepting that $\neg q$ and accepting that q. The inference is, however, valid just in case ii) is also an inconsistent set of commitments. Hale's major point now is that ii)

⁶²By failing to accept that q, the agent is not committed to accepting that $\neg q$. But he is committed to not accepting that q.

⁶³The commitment condition for negation is this: to accept that $\neg p$ is to incur a commitment to accept that $\neg p$, and not merely a commitment to not accept that p.

is not inconsistent. Not accepting that p and not accepting that $\neg p$ is the right response in cases where the available evidence leaves it unclear whether p is true or not.

A modus tollens inference should be validated by the commitment-semantic approach if this approach is a good one. The inference is, however, not validated. The approach therefore fails, or at least prima facie fails, to yield a conception of validity that, when the constituent sentences of standardly accepted patterns of inference are construed as having no truth-value, preserves the validity of these patterns of inference.

3. *Second Proposal*

A first reply to Hale's objection is that it rests on a mistaken account of what the commitment condition is for a conditional (ibid.:147). Hale has taken Blackburn's view to be that the complex commitment incurred by accepting ' $p \rightarrow q$ ' is the commitment described by (1). But perhaps Blackburn's view is that the commitment incurred is a commitment to:

(2) either accept that $\neg p$ or accept that q .

Does the latter commitment condition for a conditional validate a modus tollens inference? An agent who accepts the premises of such an inference but fails to accept its conclusion will have either of the following two sets of commitments: i) accepting that $\neg p$, accepting that $\neg q$ and not accepting that $\neg p$ or: ii) accepting that q , accepting that $\neg q$ and not accepting that $\neg p$. Both sets are inconsistent. The first set involves accepting that $\neg p$ and not accepting that $\neg p$ whereas the second set involves accepting that q and accepting that $\neg q$. The inference is therefore validated.⁶⁴

⁶⁴This commitment condition also validates a modus ponens inference. One set of commitments

This commitment condition for a conditional has, however, an unattractive feature. The argument to show this has as its premise the plausible thesis that any account of a conditional should be such that when the account is adopted, all instances of ' $p \rightarrow p$ ' are easily acceptable. If ' p ' is a sentence with a truth-value, then ' $p \rightarrow p$ ' is a tautology and if ' p ' is taken to express an attitude, it is uncontroversial, though uninformative, to assert anything of the form ' $p \rightarrow p$ '. On the first account of what is involved in accepting a conditional, ' $p \rightarrow p$ ' raises no problems. The incurred commitment to either not accept that p or accept that p is easily acceptable even in cases where there is not sufficient evidence to establish the truth-value of p . In such cases, one simply chooses to not accept that p .

On the second and stronger account of what is involved in accepting a conditional, this possibility is, however, not open. On this account, one is forced to either accept that $\neg p$ or accept that p . But in cases where the available evidence is insufficient to establish whether or not that p (in Hale's terminology, such cases are 'states of information neutral with respect to p ' or simply 'neutral information states'), we should neither accept that p nor accept that $\neg p$. This means that in cases where there is insufficient evidence to establish whether or not that p , we should be unwilling to accept ' $p \rightarrow p$ ' since this commits us to something that we do not want to be committed to. It is, however, then the case that this account of the conditional lacks a feature that any plausible account should have: namely the feature of making it unproblematic to accept ' $p \rightarrow p$ ' irrespective of what sentence ' p ' is.

There are two possibilities for rejecting this argument. Firstly, one can deny the existence of neutral information states. If one denies the existence of such states,

involves: accepting that $\neg p$, accepting that p and not accepting that q . The other set involves: accepting that q , accepting that p and not accepting that q .

then, on the stronger account of the conditional, ' $p \rightarrow p$ ' is easily acceptable for every p . Neutral information states seem, however, to be a common phenomenon. Secondly, one can reject the argument by rejecting the view that for every p , ' $p \rightarrow p$ '. If it is not the case that for every p , ' $p \rightarrow p$ ', then it cannot be a requirement on a plausible account of the conditional that it for every p , makes ' $p \rightarrow p$ ' easily acceptable. It seems, however, to be a very radical and counterintuitive move to reject that for every p , ' $p \rightarrow p$ '. To reject this would involve a denial of the law of self-entailment.

The overall conclusion of Hale's discussion is that an expressivist who tries to develop a commitment-semantics as a means to deal with the Frege-Geach point is faced with a dilemma: if he defends the weaker commitment condition for a conditional, then in cases where there is insufficient evidence to establish whether or not that p , the account fails to make valid modus tollens inferences with ' p ' as the antecedent of the conditional. If he defends the stronger commitment condition, then he offers an account of the conditional that can validate modus tollens inferences but it is an unacceptable consequence of this account that there will be instances of ' $p \rightarrow p$ ' that are not easily acceptable.

4. A Reconstruction of the Argument

In his response to Hale, Blackburn stresses the fact that when Hale ponders the question of exactly what one is committed to if one accepts ' $p \rightarrow q$ ', Hale only considers two possibilities: namely (1) and (2). Blackburn quickly dismisses the idea that (2) should be a plausible commitment condition for a conditional. Blackburn agrees with Hale that any account of the conditional should make it unproblematic to accept ' $p \rightarrow p$ ', no matter what sentence ' p ' is taken to stand for. (2) can only make

unproblematic the acceptance of 'p→p' for all 'p' if it is denied, for all p, that one can be in a state of neutral information with regards to p. Blackburn, however, contends that such a denial makes no sense: "Expressivism is perfectly hospitable to neutral information states ..." (Blackburn 2002:168).

Now, if (2) is unacceptable for the expressivist, what about (1)? As will be remembered, Hale's objection to taking (1) to be the commitment condition for a conditional is that this commitment condition fails to make valid modus tollens inferences. To make Hale's objection even more clear, Blackburn reconstructs Hale's argument and introduces some new terminology. (ibid.:169). In addition to 'accepting that p', we now have 'rejecting that p' and 'shoulder-shrugging p'. To reject that p is equivalent to accept that $\neg p$ and, similarly, to accept that p is equivalent to reject that $\neg p$. To shoulder-shrug p is neither to accept that p nor to reject that p. To shoulder-shrug p is the right attitude to adopt towards p if one is agnostic about p and/or in a state of neutral information with regards to p. If one shoulder-shrugs p, then one also shoulder-shrugs $\neg p$. It is by definition impossible to accept that p and shoulder-shrug $\neg p$. As noted, if one accepts that p, then one automatically rejects that $\neg p$, and rejection of $\neg p$ cannot coexist with shoulder-shrugging $\neg p$. Finally, Blackburn introduces the term 'avoid'. This term covers both rejection and shoulder-shrugging. One can, that is, avoid p in two distinct ways: either by rejecting it or by shoulder-shrugging it.

If this new terminology is adopted and if the commitment condition of a conditional is taken to be, not exactly (1), but a slight reformulation of it: namely:

(1*) either avoiding p or accepting that q,

then the force of Hale's objection is, according to Blackburn, clear. (1*) has the merit of making acceptance of 'p→p' unproblematic no matter what sentence p is taken to stand for. If p is a sentence such that an agent is in a neutral state of information with regards to it, then the agent just discharges the disjunctive commitment incurred by accepting 'p→p', by shoulder-shrugging p.

What about the case in which an agent accepts the premises of a modus tollens inference but fails to accept its conclusion? If the commitment condition for a conditional is the one given by (1*), then such an agent does not necessarily have an inconsistent set of commitments. Let us just consider the two scenarios in which the commitment incurred by accepting 'p→q' is resolved by 'avoiding p'. i) By accepting the minor premise of the inference, the agent rejects that q. If he avoids p by rejection, then he rejects that p. By failing to accept the conclusion, he fails to reject that p. This means that the agent rejects that p and does not reject that p. This is inconsistent: ii) By accepting the minor premise of the inference, the agent rejects that q. If he avoids p by shoulder-shrugging, then he shoulder-shrugs p. By failing to accept the conclusion, he fails to reject that p. In this scenario, there is no inconsistency. More precisely, there is no inconsistency involved in shoulder-shrugging p and failing to reject that p.

5. Blackburn's Response

Blackburn does not see himself as falling victim to Hale's objection because he is of the opinion that neither of Hale's two alternatives regarding the commitment condition of a conditional matches the one that he actually offers in *Ruling Passions*. In his book, Blackburn says that when one accepts a conditional (or a disjunction), then:

(3) "I am in a state in which if one side is closed off to me, I am to switch to the other - or withdraw" (Blackburn 1998:71).

Blackburn seeks, in other words, to avoid Hale's dilemma by arguing for two claims: a) that the choice that the expressivist faces with regard to the commitment condition of a conditional is not, as Hale seems to think, a choice between (1) and (2) but a choice between (1), (2) and (3), and b) that (3) avoids the problems that marred the other two alternatives. What are the alleged merits of (3)? Blackburn stresses that 'closed off' in (3) means 'rejected' and not merely 'avoided': "In present terms this means that if one side is rejected, I am to accept the other or withdraw" (Blackburn 2002:171).

Let us first focus on ' $p \rightarrow p$ '. This conditional is to be equated with the disjunction ' $\neg p \vee p$ '. The idea is that by accepting the conditional, an agent enters into a mental state such that if one side of the disjunction is rejected, then the other side is to be accepted. Here it is important to note that the commitment condition of the conditional is of a *conditional* form: it says that *if* one side of the disjunction is rejected, *then* the other side is to be accepted. (3) is silent on the issue of what is to happen if one side of the disjunction is shoulder-shrugged and it leaves shoulder-shrugging p an open possibility. This commitment condition works well for ' $p \rightarrow p$ '. If an agent accepts ' $p \rightarrow p$ ', then he is committed to accepting the disjunction ' $\neg p \vee p$ '. If he then rejects, say, ' $\neg p$ ', then he must accept that p or withdraw his initial acceptance.

How does (3) work in relation to a modus tollens inference? Let me try, on Blackburn's behalf, to translate the content of (3) into a form that resembles the form in which Hale has presented the various suggestions as to what the commitment

condition is for a conditional. By accepting 'p→q', an agent incurs the following commitment:

(3*) If $\neg p$ is rejected, then accept that q and if q is rejected, then accept that $\neg p$.

Where 'R' stands for 'rejected' and 'A' stands for 'accept', (3*) can be formularised like this: $[R(\neg p) \rightarrow A(q)] \ \& \ [R(q) \rightarrow A(\neg p)]$. Again, (3*) leaves shoulder-shrugging p an open possibility. Now, if (3*) is taken to be the commitment condition for a conditional, consider an agent who accepts the premises of a modus tollens inference but fails to accept its conclusion. By accepting the major premise, he incurs the commitment described in (3*). By accepting the minor premise he rejects that q. This means that he must accept that $\neg p$. The commitment condition explicitly says that if q is rejected, then $\neg p$ is to be accepted. An agent who accepts the premises of a modus tollens inference but fails to accept its conclusion therefore fails to fulfil the commitment that he has incurred by accepting the major premise of the inference.

The overall form of (3*) is, in my reconstruction, a conjunction. The commitment described by (3*) is therefore fulfilled just in case one satisfies both of the conjuncts. An agent who accepts the premises of a modus tollens inference but fails to accept its conclusion fails to fulfil the second conjunct and therefore fails to fulfil the overall commitment incurred by accepting the conditional. What about an agent who accepts the conditional, accepts that $\neg q$ and accepts that $\neg p$? Intuitively, we want to say that he has fulfilled the overall commitment. (3*) nicely secures this result. The agent rejects that q and accepts that $\neg p$. Thereby he satisfies the second conjunct. By not rejecting that $\neg p$, he satisfies the conditional of the first conjunct

and thereby satisfies this conjunct. This means, in turn, that he has fulfilled the overall commitment.

6. *A Problem For Blackburn's Response*

What should we think of Blackburn's reply? I think that Blackburn is right in believing that (3*) avoids the problems of (1) and (2). As shown above, (3*) works well for both 'p→p' and a modus tollens inference. Does this mean that Blackburn is on safe ground? Not exactly. Blackburn treats conditionals as disjunctions. This means that Blackburn will give (3) as an answer to what one is committed to if one accepts 'p v q'. One incurs a conditional commitment:

"Suppose I hold that either John is to blame, or he didn't do the deed. Then I am in state in which if one side is closed off to me, I am to switch to the other - or withdraw the commitment" (Blackburn 1998:71).

Keeping in mind that rejecting that p is equivalent to accepting that ¬p, and letting 'A' stand for 'accept', the commitment incurred by accepting 'p v q' can be formularised like this: $[A(\neg p) \rightarrow A(q)] \& [(A(\neg q) \rightarrow A(p))]$

I will now raise an objection to the view that (3) is an acceptable account of the commitment condition for disjunction. Consider the disjunction '(p&q) v (p&r)'. Let us call this disjunction 'D'. D is a particular instantiation of the logical form 'A v B'. From each of the disjuncts of D, p can be derived. This means that by accepting D, one incurs an unconditional commitment in terms of accepting a sentence. Merely by accepting D, one is committed to accepting that p. Nothing has to happen in order for this commitment to kick in. Depending on whether p expresses a proposition or expresses an attitude, one is unconditionally committed to either holding a proposition to be true or endorsing an attitude. D is a counterexample to Blackburn's

commitment-theoretic account of disjunction. This account cannot be correct since there is at least one instantiation of a disjunction that is such that when it is accepted, one incurs an unconditional commitment to accept a sentence.

Blackburn can reply to this objection by saying that it fails to take into account an important distinction. This is the distinction between: a) the claim that acceptance of a disjunction carries with it no unconditional commitment to either of the disjuncts, and b) the claim that acceptance of a disjunction carries with it no unconditional commitment at all. The proposed counterexample is only a counterexample to Blackburn's commitment-theoretic account of disjunction on the assumption that this account commits Blackburn to b). At this stage, Blackburn can, however, simply deny that it is b) that he is committed to. He can insist that depending on the exact content of the disjuncts, acceptance of a disjunction sometimes carries with it an unconditional commitment to a sentence. He could continue by saying that D is an example of a disjunction that is such that acceptance of it carries with it an unconditional commitment to a sentence that is different from its disjuncts.

However, this reply only works if Blackburn's commitment-theoretic account of disjunction enables him to explain how an unconditional commitment to a sentence can follow from acceptance of a disjunction. The major problem for Blackburn's commitment-theoretic account of disjunction is that this account does not facilitate such an explanation. The proposed account fails to validate any inference that proceeds by disjunction elimination.

Consider again D. Blackburn ought to be in a position to say that acceptance of D entails a commitment to accept that p. No inconsistency results, however, from pairing acceptance of D with failure to accept that p. Blackburn therefore has no right

to say that acceptance of D entails a commitment to accept that p. This can be brought out by considering how a proof proceeds of the inference: $A[(p \& q) \vee (p \& r)] \vdash A(p)$.

1	(1) $[A \neg(p \& q) \rightarrow A(p \& r)] \& [A \neg(p \& r) \rightarrow A(p \& q)]$	1 ASS ⁶⁵
2	(2) $\neg A(p)$	2 ASS ⁶⁶
1	(3) $A \neg(p \& q) \rightarrow A(p \& r)$	1 &E
1	(4) $A \neg(p \& r) \rightarrow A(p \& q)$	1 &E
2	(5) $\neg A(p \& r)$	2 L ⁶⁷
2	(6) $\neg A(p \& q)$	2 L
1,2	(7) $\neg A \neg(p \& q)$	3,5 MTT
1,2	(8) $\neg A \neg(p \& r)$	4,6 MTT

The proof stops here. The individual commitments of an agent who accepts the premise and fails to accept the conclusion are the ones in line (2), (5), (6), (7) and (8). This set of commitments is not inconsistent. In particular, there is nothing inconsistent about $\neg A(p \& r)$ and $\neg A \neg(p \& r)$ and there is nothing inconsistent about $\neg A(p \& q)$ and $\neg A \neg(p \& q)$. Since the set of individual commitments is not inconsistent, Blackburn cannot account for the validity of this simple inference. This is very unfortunate when you take into account that in propositional logic, the inference ' $(p \& q) \vee (p \& r) \vdash p$ ' can be shown to be valid by a simple proof involving disjunction elimination.

To this Blackburn might reply that it is a mistake to think that the proof stops at line (8). It actually continues like this:

1,2	(9) $A(p \& r)$	8 K ⁶⁸
-----	-----------------	-------------------

⁶⁵The commitment incurred by accepting the premise.

⁶⁶The commitment incurred by failing to accept the conclusion.

⁶⁷L is the rule that $\neg A(A)$ entails $\neg A(A \& B)$

⁶⁸Rule K is the rule that $\neg A(\neg A)$ entails $A(A)$.

1,2	(10) $A(p)$	9 M ⁶⁹
1,2	(11) \wedge	2, 10 CD
1	(12) $\neg\neg A(p)$	2, 11 CD
1	(13) $A(p)$	12 DN

However, as Blackburn admits himself, rule K is not acceptable. Not accepting that $\neg p$ does not commit one to accepting that p . One can, for example, be in a state of neutral information with regards to p . In such a situation, it is consistent to not accept that $\neg p$ and to not accept that p .

Another, more simple example might help to bring out Blackburn's difficulty in accounting for the validity of inferences whose validity can be shown by application of the rule of disjunction elimination. Consider the inference: $A(p \vee p) \vdash A(p)$.

This is a valid inference in propositional logic.⁷⁰ However, on Blackburn's commitment-theoretic account of disjunction, no inconsistent set of commitments follows from accepting the premise and failing to accept its conclusion.

1	(1) $[A(\neg p) \rightarrow A(p)] \ \& \ [A(\neg p) \rightarrow A(p)]$	1 ASS
2	(2) $\neg A(p)$	2 ASS
1	(3) $A(\neg p) \rightarrow A(p)$	1 &E
1,2	(4) $\neg A(\neg p)$	2,3 MTT

The proof stops here and no inconsistency has emerged. In particular, there is nothing inconsistent about $\neg A(p)$ and $\neg A(\neg p)$.

⁶⁹Rule M is the rule that $A(A\&B)$ entails $A(A)$ and $A(B)$.

⁷⁰More precisely, ' $p \vee p \vdash p$ ' is a valid inference in propositional logic.

7. Expressivism Entails Revisionism

The two above examples warrant the conclusion that Blackburn has not presented us with an acceptable commitment-theoretic account of disjunction. This is not a small and insignificant point. Let me defend this claim. The expressivist should be allowed to explain the validity of evaluative instantiations of standardly accepted patterns of inference in some other way than in terms of necessary truth-preservation. If this is not allowed, then a victory over expressivism is secured from the very outset of the debate between expressivism and its various descriptivist opponents. Such a victory is, however, not of much interest. The starting-point of a debate about what analysis to give of moral sentences should, in my view, include the observation that there are certain patterns of inference in natural language that are very compelling. One strand of the debate should then proceed by looking at how the various contenders in the debate explain this compellingness. Any theory that has no convincing explanation of this should be rejected. This approach secures a theory-neutral starting-point that allows all theories a first move.

If this methodological view is adopted, then Blackburn's alternative explanation of validity should be taken serious. What Blackburn attempts to do is exactly to tackle the question of why certain patterns of inference in natural language are compelling. The problem is, however, that Blackburn's explanation runs into difficulties. In particular, the problem is that whereas (3) has the merit of making an agent inconsistent if he flouts inferences of the form of modus ponens and modus tollens, and thereby indirectly has the merit of validating these specific patterns of inference, it does not validate *all* the standardly accepted patterns of inference.

This is a very unfortunate result for Blackburn's overall quasi-realist project. This is so because the main promise of this project is that expressivism does not

entail revisionism about everyday moral discourse. However, it now seems that revisionism about moral discourse actually is a consequence of expressivism. Certain very simple patterns of inference that are valid under a descriptivist and truth-functional account of moral sentences, come out as invalid when an expressivist account of moral sentences is in place.

This concludes my discussion of expressivism and the Frege-Geach point. Despite the fact that expressivism fares well on the issue of moral supervenience, it should nevertheless be rejected. As I have tried to show in this chapter, Blackburn fails in his most recent attempt to overcome the Frege-Geach point. This point therefore still constitutes a major obstacle to the plausibility of expressivism.

It is important to be aware that the objection in section six (assuming that it is sound) does not conclusively show the falsity of expressivism. It is a relatively technical point, and the expressivist might be able to overcome it by doing either of two things: a) hang on to the general commitment-theoretic approach, but make changes in his view about what the commitment conditions are for the various evaluative compounds. b) abandon the general commitment-theoretic approach, and develop some other strategy for dealing with the Frege-Geach point. The fact that routes a) and b) are open to the expressivist does not, in my view, render the critique of section six irrelevant and/or unimportant. Philosophical progress derives essentially from a dialectical process, and the critique offered in section six, contributes, I hope, to that process.

Chapter 8

Ethical Naturalism and the Open Question Argument

1. *Introductory*

If a methodological naturalist, who is interested in the internal accommodation project outlined in chapter one, should not be an expressivist, what then should he be? In this chapter, I begin to explore, somewhat indirectly, the realist alternative to expressivism that Cornell realism represents. In section two, I address the question of what realism and moral realism is, and I argue for a certain way of characterising these general positions. In section three and four, I discuss analytic ethical naturalism and the 'open question argument'. I argue that the latter argument is very weak, and that it does not succeed in showing the falsity of analytical ethical naturalism. In the last section, I draw attention to the fact that even if the open question argument was a sound argument, this would not be something that forces the methodological naturalist into the arms of expressivism. Analytic ethical naturalism is not the only realist alternative to expressivism.

2. *Realism*

Consider these two claims:

- a) moral discourse is as its surface-grammar suggests: namely truth-apt (this involves the claim that moral predicates have a referring semantic function).
- b) moral predicates do not refer to sui generis moral properties, but to properties that have a place within a general naturalistic world view.

The way of carrying out the internal accommodation project, that I presently want to consider, has a) and b) as its core claims. The general naturalistic world view that is assumed in this way of carrying out the internal accommodation project is similar to the kind of naturalistic world view described in chapter one. This is a rather narrow naturalistic world view and this means that what is involved in b) is ontological reductionism. Moral properties are taken to be either strictly identical to, or constituted by, natural properties. If one adopted a more expansive view about what naturalism entails in terms of ontological commitments, then it is not at all clear that ontological reductionism would be a result of b). On such a line of thought, one would reject Moore's view about the ontological status of moral properties, claim that moral predicates have a referring semantic function, and insist that, whereas the properties that moral predicates refer to, are not reducible to any of the properties that occur in the explanatory theories of our best empirical science, these properties have a place within a general (expansive) naturalistic world view.⁷¹

Features a) and b) of this way of carrying out the internal accommodation project entail a version of moral realism. The question of what exactly moral realism is, is best answered via an exposition of what realism in general is. Realism (or metaphysical realism) can be preliminary described as the view that the entities, properties and facts referred to in a certain area of discourse have objective existence. Consider discourse about the external world: our discourse about the external world contains reference to various objects, and to be a realist about this discourse is to believe that these objects have objective existence. A sentence such as 'There is a car parked in front of the church' is part of our discourse about the external world. It involves reference to two objects and the realist believes that these two objects exist

⁷¹See (McDowell 1995) for an outline and defence of this general line of thought.

objectively. Moral realism is the view that properties such as goodness, rightness and wrongness that are commonly referred to in moral discourse have objective existence, and that there are objective moral facts.

This preliminary characterisation of metaphysical realism is of course only helpful if the notion of 'objective existence' is spelled out. One way of casting light on this notion is to explain it in terms of the notion of 'mind-independence'.⁷² In connection with external objects such as cars, the notion of mind-independence can be explained as the idea that such objects exist and that their existence does not depend upon what we think about them or upon what concepts we possess. External objects have a mind-independent existence in the sense that they would exist in a world without minds.

There is, however, a problem involved in invoking the notion of 'mind-independence' in an explication of the notion of 'objective existence'. This strategy makes it impossible for the realist about mental states to make plain what his realism consists in. To be a realist about mental states is to believe that mental states have objective existence. The realist about this area of discourse cannot, however, explain what objective existence is, by saying that mental states are objective in the sense that they would exist in a world without minds. The reason for this is simply that in a world without minds, there would be no mental states. One might reply to this by saying that the fact that there is a certain area of discourse in relation to which the originally suggested definition of metaphysical realism is inadequate, should not discredit the definition. It is, one might argue, quite acceptable since it works well for so many other areas of discourse. The reply to this line of thought should be that since metaphysical realism is a position that one can adopt about any area of

⁷² See (Timmons 1999:35).

discourse, the definition of the notion should be 'area-independent'. This means that the definition should be such that no matter what area of discourse one is a realist about, one should be able to explain the 'realist' part of one's position by offering the definition.

I suggest therefore that the notion of 'ascription-independent' is substituted for the notion of 'mind-independent'. On the realist picture, mental states are objective in the sense that they exist and belong to objects independently of our ascriptions, in thought or speech, of mental states to objects. Similarly, the defining feature of moral realism would, on this approach, be the idea that moral properties are objective in the sense that they belong to objects independently of our ascriptions of moral properties to these objects.

Metaphysical realism often goes hand in hand with a certain general semantic theory. This theory says that the discourse that we are realist about is truth-apt. The sentence 'There is a car parked in front of the church' is in this sense truth-apt. The sentence is, commonly, used to assert a proposition, and propositions can be evaluated in terms of truth and falsity. This leads to the theory of truth that characteristically goes hand in hand with metaphysical realism. Declarative sentences of an area of discourse that is given a realist interpretation, express propositions, and propositions are true if and only if they correspond to the way the world is. This means that the propositions that are expressed by the declarative sentences of an area of discourse that is given a realists interpretation, are made true and false by the way the world is. Here it is important to note that whereas being a realist about a certain area of discourse entails a commitment to the view that the statements of the discourse are truth-apt, the reverse does not hold. For example, one can take moral

statements to be truth-apt without being committed to a realist interpretation of the discourse that these statements belong to.⁷³

A moral realist is a metaphysical realist about moral discourse, and he is therefore committed to the following ontological and semantic theses:

- There are (instantiated) moral properties and facts.
- Such properties and facts are objective (i.e. ascription-independent).
- Moral discourse is truth-apt.
- The statements constituting the discourse are either true or false (depending on whether or not they correspond to objectively existing moral facts).

3. *Narrow Reduction*

If a methodological naturalist is to be a moral realist, then he is faced with the task of accommodating the tenets of moral realism within the naturalistic world view. One way of carrying out this task is via what can be labelled 'narrow reduction' (Timmons 1999:37). The idea behind narrow reduction can be introduced with the use of a non-moral example. For the methodological naturalist, the statement 'Jones is punctual' is problematic. It contains reference to a property of punctuality, and this property seems ontologically dubious. Let us call such a problematical statement a 'higher level statement' (or 'A-statement') and let us call the problematical property a 'higher level property' (or 'A-property'). The methodological naturalist can carry out the accommodation project if he can reduce higher level properties to properties that fit into his naturalistic world view. Let us call such properties 'lower level properties' (or 'B-properties'). If it can be shown that the A-property of punctuality is identical to a certain B-property, then the A-statement 'Jones is punctual' is not problematic

⁷³See (Wright 1996) and (Skorupski 1999).

anymore. The property that is referred to in the statement is a property that fits into a naturalistic world view, and it is therefore unproblematic for the methodological naturalist to use and accept the statement.

The important question is, however, how this reduction is to be carried out. How is it possible to show that an A-property is identical to a B-property? According to narrow reductionism, this is possible if and only if it is possible to show that the predicates that stand for A-properties are synonymous with predicates that stand for B-properties. Applied to the example presently under consideration, this means that just in case it is possible to find a naturalistic predicate that means the same as 'punctual', are we warranted in saying that the property of punctuality is identical to the property that this naturalistic predicate refers to. The methodological naturalist might suggest that 'punctual' means the same as 'always arriving on time'. This latter predicate refers to the property of always arriving on time, and if this piece of conceptual analysis is accepted, then the methodological naturalist is warranted in saying that the property of being punctual is identical to the property of always arriving on time.

Methodological naturalists who try to carry out the accommodation project via narrow reduction can be called 'analytic naturalists'. They subscribe to a synonymy criterion of property identity. According to this criterion, the referent of an A-term is identical to the referent of a B-term if and only if, the two terms are synonymous. If we apply the synonymy criterion of property identity to moral discourse, then we get 'analytic ethical naturalism'. In the statement 'Lying is wrong', the predicate 'wrong' occurs. This predicate refers to the property of wrongness. In order to be able to identify this property with a natural property, the analytic ethical naturalist must be able to find a predicate that means the same as 'wrong'. He might

suggest that the predicate 'diminishes overall well-being' means the same as 'wrong'. Just in case, the analytic ethical naturalist is successful in establishing this claim, is he warranted in holding that wrongness is identical to diminishing overall well-being.

The analytic ethical naturalist can now say that since the predicate 'wrong' refers to a certain natural property, there is nothing problematic in him using and accepting sentences in which this predicate occur. The use and acceptance of such sentences does not commit him to the existence of a sui generis property of wrongness. The analytic ethical naturalist can furthermore accept the claim that there is an objective property of wrongness in the world (this is why his position is a version moral realism). He can accept this claim because the property of wrongness is identical to a certain objective, natural property.

From this account of analytic ethical naturalism it is easy to see that it is only going to work if moral terms have analytically true naturalistic definitions. Defenders of analytic ethical naturalism of course believe that such definitions can be found, and they therefore subscribe to a thesis that can be named: 'analytic semantic naturalism' (ASN). According to ASN, fundamental moral terms like 'good' have analytically true naturalistic definitions. Since analytic ethical naturalism is only going to work if ASN is true, the plausibility of this thesis must be assessed in order to answer the question of whether or not analytic ethical naturalism is a plausible metaethical theory.

4. The Open Question Argument

One particular argument looms large in the discussion of ASN. This is Moore's 'open question argument' (OQA).⁷⁴ As a prelude to this argument, a distinction should be

⁷⁴ See (Moore 1993:chapter one).

made between two different construals of ethical naturalism. Ethical naturalism can be construed as a thesis about the meaning of words. Ethical naturalism is here equivalent to ASN. Ethical naturalism can, however, also be construed as a thesis about the ontological status of the property of goodness. Ethical naturalism is here an ontological doctrine that essentially involves the claim that the property of goodness is identical to some other property, such as, say, pleasantness. Ethical naturalism (in both of these construals) is the intended target of the OQA.

In order to show that the property of goodness cannot be reduced to any other property, Moore puts forward the OQA. The argument claims that if ethical naturalism (understood as an ontological doctrine) is true, then a certain question is not an open question. This question is, however, an open question. Therefore it cannot be the case that ethical naturalism is true.

Imagine that a certain analysis of 'good' is put forward. According to this analysis 'good' means 'N' (where 'N' is some naturalistic predicate). If this analysis is true, then the question 'Is x which is N, good?' turns into the question 'Is x which is N, N?'. This is not an open question that it, according to Moore makes sense to ask.⁷⁵ The question is a closed question to which the answer is trivially 'yes'. Moore insists, however, that the question 'Is x which is N, good?' is an open question that it always makes sense to ask, no matter what naturalistic (or non-naturalistic predicate) 'N' is taken to stand for. Therefore it cannot be the case that 'good' means the same as 'N' (Moore 1993:67). This view of Moore's can also be expressed by saying that 'good' lacks naturalistic (or non-naturalistic) synonyms.

⁷⁵ Here, Moore's way formulating himself is rather misleading. Strictly speaking, it is not the case that it makes no sense to ask this question. It actually makes perfect sense to ask it, but what is special about the question is, that it is a question to which the answer is trivially 'yes'.

A traditional objection to Moore's argument consists in the claim that it does not prove what Moore takes it to prove.⁷⁶ The OQA is a semantic argument. It is an argument which is concerned with the meaning of different words. The argument may prove that 'good' does not have the same sense as any other predicate. Moore's argument may, that is, prove that ASN is false. But from this, it does not follow that the property of goodness is a *sui generis* property that is different from any other property. Moore's argument does not show that ethical naturalism, understood as an ontological doctrine, is false. The falsity of the first kind of naturalism does not entail the falsity of the second kind of naturalism. This can be seen by considering what the result would be if Moore's way of reasoning was sound. If it were, this would have the paradoxical result of showing that e.g. water is not identical to H₂O. 'Water' does not mean the same as 'H₂O'. As a result of this, the question 'Is x which is water, H₂O?' is an open question. It is not a question to which the answer is trivially 'yes' because 'water' and 'H₂O' have the same sense. Now, if Moore's way of reasoning is sound, then we can go on to conclude that a sample of water is an entity that is ontologically different from a sample of H₂O molecules. But this is plainly false.

Is it possible to explain why Moore makes this unwarranted slide from a semantic conclusion about the meaning of 'good' to an ontological conclusion about the property of goodness? If we focus on the theory of meaning to which Moore subscribes, then such an explanation can be found. Moore is of the opinion that names and predicates denote certain objects or properties. Names and predicates get, moreover, their meaning from the object or property that they denote. If a name or predicate does not denote a certain object or property, then the name or predicate is without meaning. In line with this theory, the predicate 'good' gets its meaning from

⁷⁶ See (Pigden 1993:426).

the property it denotes. Moore takes it to be the case that 'good' denotes a simple and indefinable property of goodness. Other predicates have the same structure as 'good'. 'Yellow' denotes the property of yellowness and 'pleasant' denotes the property of pleasantness. Since predicates get their meaning from the properties that they denote, a straight forward conclusion can be drawn from the fact that 'good' is a meaningful predicate that does not mean the same as any other predicate: namely that the property of goodness is different from any other property. In short, Moore subscribes to the synonymy criterion of property identity. But as the above examples illustrate, this criterion is clearly false. A possibility that Moore does not envisage consists in the idea that two non-synonymous predicates may denote the same property.

As it is evident from the above discussion, the OQA does not succeed in showing that ethical naturalism (understood as an ontological doctrine) is false. Does it even succeed in showing that ASN is false? The OQA rests on the assumption that a correct analysis of a concept cannot be unobvious and informative to a competent speaker (a speaker who knows the meaning of the analysandum and the analysans). To see this, assume for a moment the opposite: assume, that is, that an analysis of concept C can be given in terms of concept C*, and that this analysis is unobvious and informative to an agent who is a competent speaker in the above sense. If this assumption is true, then the analytical naturalist can give the following, convincing answer to Moore's objection. Due to the truth of the above assumption, it is quite possible for, some analysandum (C) and analysans (C*), that the statement 'x which is C is C*' is unobvious and informative for a competent speaker. It is therefore quite possible that for such a speaker, the question 'is x which is C, C*?' is an open question in the sense that it is not a question to which the answer is trivially 'yes'. For the speaker, this question is not like the question 'is x which is C, C?' which

obviously is not an open question in the present sense of 'open'. The point that the analytical ethical naturalist can now make is, that the fact that the question 'is x which is C, C*?' is open for a competent speaker does not necessarily show that C* does not have a sense equivalent to the sense of C. Because of the truth of the assumption that there can be unobvious and informative analyses, the fact that the question 'is x which is C, C*?' is open for a competent speaker is compatible with it being the case that the analysis of C in terms of C* is a correct analysis.

To guard against this move, Moore, and others who want to invoke the OQA against ASN, has to insist that there cannot be such a thing as unobvious and informative analyses. This means that there is only one way of answering the question of whether the OQA proves the falsity of ASN, and that is to establish the truth-value of the assumption that there can be unobvious and informative analyses. The assumption appears to me to be true. For example, consider the concept of a triangle (C), and the concept of a three-sided, closed geometrical figure with an angle-sum of 180 degrees (C*). An agent can, it seems to me, be a competent speaker, and yet find the correct analysis of C in terms of C*, unobvious and informative. Why is this possible? One reason has to do with what it is to be a competent user of concept C. It seems to me that one is a competent user of this concept if and only if one applies it to the correct class of objects, that is to the class of triangles. But one can be a competent user of C in this sense, and yet not be fully conscious of the criteria, or sense, by which one applies the concept. In this case, what the analysis of C in terms of C* does, is to make explicit the sense of C, and for someone who has not previously been fully conscious of the sense of C, this analysis of the concept, is likely to be unobvious and informative. The analysis makes something explicit to the agent that he has hitherto only been implicitly aware of.

So, it should be concluded that there can be unobvious and informative analyses.⁷⁷ Moore's argument is therefore not successful in showing that ASN is false. The assumption that the OQA rests on is false, so as the argument stands, the analytical ethical naturalist has nothing to fear from the OQA. The fact that 'is x which is N, good?' is an open question, does not automatically show the falsity of an analysis of the concept of being good in terms of being N. For all the openness of this question, 'N' might be synonymous with 'good'.

It is important to note that these considerations do not vindicate analytical ethical naturalism. The main claim of this position is that fundamental moral terms can be given an analysis in naturalistic vocabulary. If the analytical ethical naturalist adds to this, the claim that these analyses are going to be unobvious and informative, then he has a convincing reply to the OQA. This of course still leaves the analytical ethical naturalist with the substantial task of providing us with convincing examples of such analyses. The mere fact that there can be unobvious and informative analyses does not in itself show that there actually are unobvious and informative naturalistic analyses of fundamental moral terms.

5. Another Alternative

Two main points emerge from the discussion of ethical naturalism and the OQA. Firstly, that the OQA cannot be used to show that analytical ethical naturalism is a flawed position. If the only argument against this position is the OQA, then the analytical ethical naturalist is warranted in holding that his position is an open option for a methodological naturalist who is interested in giving a realist interpretation of moral discourse.⁷⁸ Secondly, and more important for my purposes, is the point that

⁷⁷For a similar conclusion, see (Smith 1994:38) and (Pigden 1993:427).

⁷⁸For contemporary attempts to vindicate analytic ethical naturalism, see (Smith 1994) and (Jackson

even if the OQA (or any other argument) is successful in showing that ASN is false (and thereby refuting analytical ethical naturalism), then this is not something that shows that the methodological naturalist who is interested in accommodating moral discourse is forced to embrace some non-descriptive (e.g. expressivist) interpretation of moral discourse. This is so, because analytical ethical naturalism is not the only option available for a methodological naturalist who is attracted to a realist interpretation of moral discourse. Cornell realism is another alternative for such a naturalist.

Chapter 9

The Semantic Views of Cornell Realism

1. *Introductory*

Presenting a full account of Cornell realism, and the way in which the accommodation project is carried out within this tradition is a complex task. What I intend to do in this chapter is to focus on one aspect of the general program: namely its semantic tenets. The semantic views of Cornell realism have been developed in greatest detail by Richard Boyd.⁷⁹ In sections two to four of this chapter, I give an account of Boyd's position, and present a very influential argument against it, set forward by Mark Timmons.⁸⁰ In section five, I discuss Timmons' argument, and develop a defensive strategy against it. The conclusion of the chapter is that Timmons' argument does not succeed in its intended purpose: i.e. it cannot be invoked to show that the semantic views associated with Cornell realism are mistaken. This does not amount to a vindication of these views. However, the conclusion of the chapter is, if true, interesting because it means that if the semantic views associated with Cornell realism are mistaken, this is not for the reason that many writers suppose.

⁷⁹(Boyd 1988).

⁸⁰(Timmons (1999). The argument has previously been presented in a series of co-authored articles. See (Horgan & Timmons 1991), (1992a) and (1992b). For a discussion of the argument, see, for example, (Gampel 1997), (Copp 2000), (Brink 2001), (Bloomfield 2003) and (Miller 2003). Miller (2003:168) suggests that the argument poses a challenge to defenders of Cornell realism whereas Bloomfield (2003:199) concludes that the argument is fatal for Cornell realism. Gampel, Copp and Brink all argue that there are ways for the Cornell realist to overcome the argument. Their replies to the argument differ, however, from the one I shall go on to suggest.

2. *A Causal-Functional Account of Moral Terms*

Prior to presenting his argument, Timmons recapitulates Boyd's semantic views and in the process of doing so, he introduces a number of technical terms and definitions that are not found in Boyd's text. Since my aim is not to present an exegesis of Boyd's views, but rather to offer an assessment of the strength of Timmons' objection, I find it convenient in the following to account for Boyd's position by partly relying on Timmons' exposition of it.

According to Timmons, Boyd is committed to a causal regulation thesis (CRT) for moral terms.

CRT: "For each moral term *t* (e.g., 'good'), there is a natural property *N*, such that *N* and *N* alone regulates the use of *t* by humans" (Timmons 1999:58).

CRT expresses an empirical thesis to the effect that humans only use the predicate 'good' about actions, policies and institutions that possess natural property *N*. The possession of *N* by some act, policy or institution causes humans to apply 'good' to it. Timmons then says that Boyd is of the opinion that the truth of CRT allows one to treat moral terms as semantically akin to natural kind terms: that is, moral terms rigidly designate certain natural properties, and they possess synthetic definitions (ibid.:58). The semantics of Cornell realism (in Boyd's development) can be summarised in a thesis called 'Causal Semantic Naturalism' (CSN).

CSN: "Each moral term *t* rigidly designates the natural property *N* that uniquely causally regulates the use of *t* by humans" (ibid.:58).

Assume that it is the natural property *N* that causally regulates the use of 'good' by humans. According to CSN, it is then the case that 'good' rigidly designates this natural property. Not only in the actual world, but in all the possible world in which

'good' has a referring function, 'good' refers to N. Since Cornell realism, according to Timmons, sees moral properties as functional properties, the semantic view of this position can ultimately be expressed in a thesis called 'Causal Semantic Functionalism Thesis' (CSFT).

(CSFT): "Each moral term *t* rigidly designates the unique functional property that causally regulates the actual-world uses of *t* by humans" (ibid.:58).

The basic idea behind functionalism about moral properties is that such properties have an essence that is to be explained in terms of the causal relation(s) that these properties enter into. Functionalism can be further specified by the claim that moral properties are identical to second order properties, and not the first order properties that realise them.

How do we find out what the functional essence of, say, goodness is? This is largely an empirical matter, just as it is an empirical matter what the essence of, say, water is. Let *T* be a normative theory and let '*t*' be a property-expressing moral term. For each '*t*', *T* specifies the *t*-role: the functional essence of the property expressed by '*t*'. Let *T_c* be, say, hedonistic utilitarianism and let '*t*' be, say, 'right'. Then *T_c* might specify the functional essence of rightness like this:

Rightness *T_c*: *x* is good iff *x* causes a maximisation of human pleasure.

T_c also specifies the *t*-role for the properties expressed by '*t**', '*t***' etc. Different moral theories give different accounts of the functional essence of rightness: i.e. they give different specifications of the *t*-role of '*t*'. Assume now that an empirical investigation is initiated that aims at finding out what the underlying nature is of the things that people apply the term '*t*' ('right') to. Assume, moreover, that it turns out that people apply 'right' to things that cause a maximisation of human well-being

(assume, that is, that human uses of 'right' is causally regulated by things that cause a maximisation of human well-being). It also turns out, we may assume, that for all the other moral terms ('t*', 't**' etc.), people apply these terms to things whose underlying nature is as is specified by Tc. Tc is then a moral theory whose claims about the functional essence of t, t*, t**, etc. correctly captures the functional essence of the properties that causally regulate human uses of 't', 't*', 't**', etc. In this case, Tc is vindicated as the correct moral theory, and the functional essence of rightness is the t-role of 't' as specified by Tc.

"The idea is that even though people often disagree morally, and even though mistaken moral beliefs commonly occur, nevertheless, human moral terms and moral beliefs still are causally regulated by some unique set of functional properties whose essence is captured by theory T. Coherentist methodology in ethics is supposedly an epistemologically appropriate way to learn about T empirically" (Horgan & Timmons 1992b:242).

So according to Timmons, Cornell realism gives a causal-functional account of the semantics of moral terms. On this account, moral terms such as 'good' function like natural kind terms in the sense that, through their use of them, ordinary competent speakers purport to refer to some objective natural (functional) property that regulates their use and whose essence is discoverable through moral inquiry (Timmons 1999:59). Timmons rejects this account of how moral terms function: he is of the opinion that there is a radical difference between the semantics of natural kind terms and moral terms.

3. Putnam's Twin-Earth Example

In his attempt to undermine the causal-functional account of moral terms, Timmons first considers how philosophers have argued for causal accounts of natural kind terms. Putnam's twin-earth thought experiment is here the important argument.⁸¹

Consider an agent who grants that human uses of 'water' is causally regulated by a certain physico-chemical substance but rejects the claim that 'water' rigidly designates the substance that happens to fill the 'water-role' on earth. Putnam subscribes to the distinction between a description being meaning-fixing and reference-fixing. The description 'clear, odourless liquid that falls from the sky....etc.', does not give the meaning of 'water'. It fixes, however, the reference of 'water' via a clause like this:

Water (earth): x is water iff x has the same molecular structure as the substance that is actually dominantly causally responsible for earthlings' perception of a clear, odourless liquid that falls from the sky...etc.

On earth, it is H₂O that fills the water-role since it is H₂O that is dominantly causally responsible for earthlings perceptions of a clear, odourless liquid....etc. This point can also be put by saying that on earth it is H₂O that plays the appropriate causal role of water since it is H₂O that is dominantly causally responsible for earthlings' perception of a clear, odourless liquid that falls from the sky...etc.

Putnam's thought experiment is designed to convince the sceptic that 'water' is a rigid designator. The thought experiment goes like this: imagine a twin-earth that is completely identical to earth except for the fact that the substance that fills the water-role on twin-earth has a molecular structure that can be abbreviated XYZ (where having a XYZ molecular structure is not identical to having a H₂O molecular

⁸¹ See (Putnam 1975a).

structure). This substance has all the superficial properties that water has on earth and the people on twin-earth (twin-earthlings) use the term 'water' to refer to it. The idea now is that twin-earth might constitute an example of a possible world in which 'water' refers to a substance that is not H₂O. On earth, 'water' refers to H₂O and on twin-earth, 'water' refers to XYZ. If this is true, then it cannot be the case that 'water' is a rigid designator.

Putnam, however, rejects the idea that twin-earth constitutes an example of a possible world in which 'water' has a referring function, but in which 'water' does not refer to H₂O. His argument is this: suppose that an earthling (E) travels to twin-earth and meets a twin-earthling (TE). As observers to this meeting, we are aware of the difference in what causally regulates earthlings' and twin-earthlings' use of 'water'. We are aware, that is, of the fact that on earth, the use of 'water' is causally regulated by H₂O whereas the use of 'water' on twin-earth is causally regulated by XYZ. Suppose now that (E) and (TE) have a conversation during which (E) says 'Ø is water' and (TE) says 'Ø is not water'.

What intuitions are generated by reflection on this scenario? According to Putnam, the intuition generated is that there is no genuine disagreement between (E) and (TE) about the nature of Ø. Our intuition is that the two parties are speaking past each other since (E) uses 'water' to refer to H₂O and (TE) uses 'water' to refer to XYZ. On Putnam's view, the situation involving (E) and (TE) is much similar to a situation in which we observe two persons talking about where the exact location is of the referent of 'Tony'. We are aware that one of them uses 'Tony' to refer to the PM of Britain whereas the other uses 'Tony' to refer to an American jazz-singer. In a situation in which one of them says that Tony is in London and the other one says that Tony is not in London, we would not at all be tempted to think that there is a

genuine disagreement between them about the exact location of Tony. We would, on the contrary, think that the two persons are speaking past each other.

Putnam's argument now proceeds by asking what semantic account of 'water' can explain the intuition created by reflection on the twin-earth example. His claim is that this intuition is explained by the truth of a certain theory of meaning. This theory is a referential theory of meaning that equates the meaning of natural kind terms with the reference of these terms. Natural kind terms do not have an intension in the descriptivist sense of 'intension'. Natural kind terms therefore cannot be given analytic definitions that reveal the intension of the terms. The meaning of a natural kind term like 'water' is a certain natural kind whose essence can be discovered through empirical investigation. If this referential theory of meaning is true, then we can see how it can be the case that the twin-earthian term 'water' does not mean the same as the earthian term 'water'. The two terms have different meanings because they have different referents. The twin-earthian term 'water' refers to XYZ whereas the earthian term 'water' refers to H₂O.

We already know that the reference of the earthian term 'water' is fixed via a clause that involves a description of the causal-functional role of water. The reference of the twin-earthian term 'water' is fixed via a much similar clause:

Water (twin-earth): x is water iff x has the same molecular structure as the substance that is actually dominantly causally responsible for twin-earthlings' perception of a clear, odourless liquid that falls from the sky...etc.

On twin-earth, it is XYZ that plays the appropriate causal role of water. So, on Putnam's view, both 'water(E)' and 'water(TE)' are rigid designators. In all possible worlds in which 'water(TE)' has a referring function, it refers to XYZ. We now have

an explanation of the intuition that (E) and (TE) talk past each other when (E) says that \emptyset is water and (TE) says that \emptyset is not water. A genuine disagreement between (E) and (TE) regarding the nature of \emptyset presupposes that both parties in the dispute mean the same with the term 'water'. (E) and (TE) do not, however, mean the same with 'water'. The referential theory of meaning is true, and on earth, 'water' refers to H₂O. This means that 'water' in the mouth of (E) means H₂O. On twin-earth, 'water' refers to XYZ. This means that 'water' in the mouth of (TE) means XYZ.

4. *Timmons' Argument*

Let us now return to the causal-functional theory of moral terms and consider Timmons' objection to it. Timmons argues that if 'good' is semantically akin to 'water' in the sense that 'good' rigidly designates the natural property that causally regulates the use of it by humans, then it should be possible to construct a twin-earth scenario with 'good' that has the feature that reflection upon it generates intuitions similar to those generated by reflection on Putnam's original scenario. Conversely, if the intuitions generated by reflection on the twin-earth scenario with 'good' are radically different from the ones generated by reflection on the original scenario, this will be empirical evidence that 'good' is not semantically akin to 'water' (Timmons 1999:60).

Suppose that 'good' functions semantically as the causal-functional theory says it does. Suppose moreover that the natural property that causally regulates human uses of 'good' is natural property N (the functional property of causing a maximisation of human pleasure). Now, consider moral twin-earth which is almost completely identical to earth. Moral twin-earthlings have a moral vocabulary that works much like human moral vocabulary. They use the terms 'good', 'right' and 'wrong' to evaluate actions, institutions and policies. The uses of these terms on

moral twin-earth have all the formal marks that we take to characterise moral vocabulary and moral practice. Twin-earthlings are, for example, generally motivated to perform acts that they take to be good. For twin-earthlings, considerations about the goodness of an act generally override other kinds of considerations, such as considerations about whether or not it is prudential from a first order perspective to perform the act. As a result of this, if some earthlings visited moral twin-earth, then they would be strongly inclined to say that the twin-earthlings' terms 'good', 'right' and 'wrong' mean the same as our terms 'good', 'right' and 'wrong'.

Let us now suppose that we investigate twin-English moral discourse and discover that the twin-earthlings' use of moral terms is causally regulated by natural properties different from the natural properties that causally regulate human uses of moral terms. The twin-English term 'good' tracks a certain natural (functional) property, but this property is, say, the natural property N^* (where N is not identical to N^*). So we may suppose that there is a difference between earth and twin-earth regarding what natural property causally regulates the use of 'good'. Such a difference implies that there is a difference between earth and twin-earth in regards to what first order moral theory is in place. On twin-earth it is, we may assume, a version of deontology that is in place. This theory specifies the functional role of goodness (roughly) as follows:

Goodness Td: x is good iff the performance or endorsement of x is caused by
a motive to adhere to the prescriptions of the categorical
imperative.

If we reflect upon this moral twin-earth scenario, what do we want to say about the meaning of the earthian term 'good' and the twin-earthian term 'good'? Two options are, according to Timmons, available (ibid.:62). On the one hand, we can say that the

earthian term 'good' does not mean the same as the twin-earthian term 'good'. This is the view we are forced to hold if we believe that moral terms are semantically akin to natural kind terms. On earth, 'good' refers to a certain natural property and on twin-earth, 'good' refers to a different natural property. The two terms have different referents and since the meaning of moral terms (*if* they are semantically akin to natural kind terms) is equivalent to their referent, then the conclusion must be that the two terms have different meanings. The earthian term 'good' is not translatable with the twin-earthian term 'good'.

According to Timmons, it is, however, the case that if we take this line, then it seems that we are unable to account for what appears to be a genuine moral disagreement between an earthling and a twin-earthling that arises if the two meet and (E) says that \emptyset is good and (TE) says that \emptyset is not good. On Timmons' view, this exchange of statements strongly suggests that (E) and (TE) have a genuine moral disagreement. What (TE) says seems to contradict what (E) says and therefore they have a disagreement. But actually there is no disagreement between them. If 'good' in the mouth of (TE) does not mean the same as 'good' in the mouth of (E), then his statement is not a contradiction of (E)'s statement.

On the other hand, we can say that the earthian term 'good' does not differ in meaning from the twin-earthian term 'good'. If we take this line, then we reject the idea that the referential theory of meaning is true of moral terms. If it is the case that the earthian term 'good' does not have the same referent as the twin-earthian term 'good' and if the two terms mean the same, then it cannot be true that the meaning of 'good' is equivalent to its referent. If this was true, then the earthian 'good' and the twin-earthian 'good' would differ in meaning. If we reject the idea that the referential theory of meaning is true of moral terms, then we are forced to reject the idea that

moral terms are semantically akin to natural kind terms and then we reject Boyd's account of the semantics of moral terms.

A consequence of saying that the earthian term 'good' and the twin-earthian term 'good' does not differ in meaning is that we will be able to say that (E) and (TE) really are engaged in a genuine moral disagreement. Since (TE) means by 'good' exactly what (E) means by 'good', then his statement is a contradiction of what (E) says.

According to Timmons, the latter option constitutes the most natural way of thinking when presented with the moral twin-earth scenario. Having considered this scenario, the intuitively natural thing to say is that the earthian term 'good' is translatable with the twin-earthian term 'good'. But if the causal-functional account of moral terms was right and moral terms were semantically akin to natural kind terms, then it should seem intuitively natural to say that the two terms differ in meaning. This intuition we do not have, according to Timmons, and this is strong empirical evidence for the claim that Boyd's account of the semantics of moral terms is wrong.

5. Replies to Timmons' Argument

What should the reply be to Timmons' argument? In my opinion it has a certain *prima facie* appeal. There are, however, a number of strategies that a defender of the causal-functional account can employ. Firstly, he could reject the empirical thesis upon which Timmons' argument rests: namely the thesis that reflection upon the moral twin-earth example generates the intuition that (E) and (TE) are engaged in a substantial moral disagreement. The truth of this thesis is not evident and since the thesis plays an important role in Timmons' argument, the naturalist could say that the

onus is on Timmons to provide evidence for its truth. Lacking any such evidence, the argument is inconclusive and suggestive.⁸²

Since I am interested in providing a defensive strategy against the strongest possible version of Timmons' argument, I am, however, going to assume the truth of the thesis upon which his argument rests. Granting that we have the intuition that Timmons says we have, a distinction can be made between two versions of his argument.⁸³ The first version rests not only on the assumption that we have the intuition in question, but also on the additional assumption that this is the correct intuition to have. But to assume that the intuition that (E) and (TE) mean the same by 'good' is the correct intuition, is to beg the question against the causal-functional account of moral terms. It is to presuppose the truth of a theory of meaning of moral terms that is non-referential (e.g. an expressivist theory). So, I contend that this version of the argument can just be rejected.

The second version of the argument does not rest upon the additional assumption. It rests only on the assumption that people actually have the intuition that Timmons mentions. This second version is not question-begging, but it is a much weaker version than the first. This version of the argument merely poses a challenge to someone defending a causal-functional account of moral terms to come up with an explanation as to why people have this intuition. Can such an explanation be provided? I think so.

What is our present situation in relation to the question of what causally regulates the use of 'good'? It seems fair to say that we do not know what property it is that plays this regulative role, and we do not even know *if* there is a unique (presumably functional) property that plays this role. It is important to note that

⁸²This strategy is noted by Copp (2000:119).

⁸³This is a point made by Miller (2003:167).

Timmons builds into his twin-earth example that we know what property it is that uniquely, causally regulates the use of 'good' on earth. This means that by asking us to consider his twin-earth example, he asks us to do two different things: firstly, to imagine ourselves being in a situation in which we know something that we at present do not know. Secondly, to consider what our intuitions would be if we were in that position.

My suggestion is that due to our present situation in relation to 'good', it is difficult for us, through the workings of our imagination, to get a proper grasp of what it would be like to be in a position in which one had the knowledge stipulated by Timmons. If the situation described by Timmons were one that were only marginally different from our present situation, then it would be relatively easy for us, through the workings of our imagination, to get an accurate picture of what it would be like to be in that position. The situation described by Timmons is, however, a situation that is significantly different from our present one, and therefore there is a greater likelihood that our imagination gives us a distorted picture of what it would be like to be in that position. This idea should appear uncontroversial. It is, for example, very probable that the picture given to me by my imagination about what it would be like to have six fingers on my left hand, is more accurate than the picture given to me about what it would be like to have two left hands.

Because it is difficult for us to get a proper grasp of what it would be like to be in a position in which one knows that the use of 'good' is causally regulated by natural property N, our judgement about what our intuition would be in that situation should be treated with some scepticism. It is quite likely that our judgement about what our intuition would be in a situation in which we have the knowledge specified by Timmons, actually is a judgement that is informed and influenced by our present

situation in which we do not have this knowledge. And a judgement that is so informed, is likely to be a judgement to the effect that our intuition is that (E) and (TE) are engaged in a substantial moral disagreement. Why is this? This is due to the fact that at present it does not appear to us that the use of 'good' tracks any unique property. On the contrary, it seems as if 'good' is applied to a number of different things that do not have any unifying, underlying property in common. In such a situation, reflection upon the conversation between (E) and (TE) is likely to generate the intuition that the two parties mean the same by 'good', but simply have different moral standards.

Is it plausible to suggest, as I have done, that there is a causal relation between our present situation in relation to 'good' and our judgement about what our intuition is about a certain twin-earth example? One way to evaluate the suggestion involves getting clear about what its implications are, and then inquiring into whether these implications are plausible. If it turns out that the implications are implausible, then this is evidence for the claim that the initial suggestion is implausible. Conversely, if the implications are not implausible, then this does something to bolster the initial suggestion.

So, what are the implications of my suggestion? One important implication is that if a change were to occur in our situation in relation to 'good', then a change would occur in our judgement about the intuition in question. Why is this? Well, to suggest that A causally influences B would look very implausible in light of evidence to the effect that B stays the same despite major changes in A.⁸⁴ We would, for example, grow increasingly sceptical about the suggestion that the amount of time spent outdoors by an individual, causally influences the individual's experienced

⁸⁴This is so only on the assumption that B is not causally over-determined: i.e. that B is not causally influenced by anything else than A.

degree of psychological well-being, if it turned out that the degree of experienced psychological well-being remained stable across big variations in the amount of time spent outdoors. If the suggestion were anyway near the truth, we would expect changes in the amount of time spent outdoors to result in changes in experienced psychological well-being.

Now, in my view, it is not implausible to think that if a change were to occur in our situation in relation to 'good', then a change would occur in our judgement about what our intuition is about the twin-earth example in question. Imagine a future state of affairs (FSA) in which we do not only know that the use of 'good' is causally regulated by natural property N, but in which we are also very comfortable with this knowledge. To be 'comfortable' in this respect means that we have had the knowledge for generations and that the knowledge is common knowledge. The vast majority of people in (FSA) have grown up being told that the use of 'good' tracks natural property N and that goodness is always co-instantiated with N-ness in acts, policies and institutions. Everybody knows that when an agent talks about goodness, he talks about N-ness. Due to these stipulations about (FSA), our situation in that state of affairs in relation to 'good' is very similar to our present situation in relation to 'water'.

Consider now what our intuition would be about Timmons' twin-earth example if we reflected upon it from the perspective of (FSA). It is not evident what the answer would be to this question. First of all, people have different intuitions. Secondly, it is, as I have tried to argue, difficult to get a proper grasp of what it would be like to be in such an unfamiliar situation as (FSA) arguably is. Some people might mistake an intuition arising from their present situation with an intuition arising from (FSA). Other people might be able to form a proper picture of what it

would be like to be in (FSA) and therefore be able to report what their intuition would be in this situation. However, my view is that there is reason to expect that the intuition in (FSA) would be different from our intuition in our present situation. Keep in mind that in (FSA) we are comfortable with the idea that when (E) talks about goodness, he talks about N-ness, and we know that when (TE) talks about goodness, he talks about N*-ness. If anything surprises us about (TE)'s use of 'good', it is not likely to be the fact that he uses it to refer to a natural property. It is, I think, more likely to be the fact that he uses it to refer to the specific natural property N* and not N.

In (FSA), 'good' is used by speakers in a way very similar to the way in which 'water' is used by present speakers. This warrants, in my view, the expectation that the intuition we will have in (FSA) about a twin-earth example involving 'good' is going to be similar to the intuition that we presently have about a twin-earth example involving 'water'. In short, it is to be expected that we would have the intuition that (E) and (TE) are not engaged in a substantial moral disagreement.

Where does all of this leave us? The answer is that a defender of causal-functional account of moral terms has nothing to fear from Timmons' moral twin-earth argument. At best this argument poses an explanatory challenge, but as I have tried to show in the previous section, this challenge can be met. From the fact that Timmons' specific argument does not show the falsity of the causal-functional account, it cannot, however, be concluded that this account is correct. There might be other arguments that bring out difficulties with this account. Any such argument deserves, however, its own attention, and my aim in this chapter has only been that of defending the causal-functional account against a single, very influential argument.

Chapter 10

Cornell Realism and Moral Supervenience

1. *Introductory*

My aim in this chapter is twofold: to give an account of some of the key metaphysical commitments of Cornell realism, and to discuss whether or not Cornell realism can give a convincing explanation of moral supervenience. In pursuing this aim, I shall concentrate on the works of David Brink (Brink 1984; 1989; 2001) since it is in his works that the most explicit formulation is found of the metaphysical commitments of Cornell realism. In section two, I outline and discuss Brink's view of moral properties. Section three explicates Brink's notions of strong and weak supervenience. In sections four and five, I discuss Brink's explanation of moral supervenience. I argue that his functionalist view of moral properties means that the explanation of moral supervenience that he explicitly offers, is not completely satisfactory. I also make explicit the exact nature of the explanatory challenge that I think Brink faces. In section six, I argue that Brink can meet this explanatory challenge. In particular, I argue that the special modal force with which moral supervenience holds is not a problem for Brink. The overall conclusion of the chapter is given in section seven, and it is that Cornell realism does not have a problem on the issue of moral supervenience.

2. Moral Properties as Natural Properties

Brink asserts that what is distinctive of ethical naturalism is the claim that moral properties are natural properties: "The ethical naturalist claims that moral facts and properties are natural (i.e. natural and social scientific) properties" (Brink 1989:156).

This general characterisation of ethical naturalism lends itself to two distinct interpretations:

- a) moral properties are identical to natural properties (the 'is' of identity)
- b) moral properties are constituted by natural properties (the 'is' of constitution)

What does it mean that moral properties are identical to natural properties? It means, or should be taken to mean, that moral predicates and certain naturalistic predicates designate the same properties in all possible worlds:

"..but as the claim that moral terms and certain natural and social scientific terms designate or express the same properties (i.e. refer to the same properties in all possible worlds" (Brink 1989:157).

To take a non-moral example: the claim that the property of being water is identical to the property of being H₂O does not mean that there is one property of being water, and another, distinct, property of being H₂O, but that these two properties are nonetheless identical. It simply means that 'water' and 'H₂O' designate the same substance in all possible worlds.⁸⁵

When construed as a theory that makes identity claims, ethical naturalism implies that moral and natural properties are necessarily identical (ibid.:157).

According to Brink, naturalistic identity claims in ethics should be construed on the

⁸⁵ The clause of necessary co-reference is important here. Without it, one would, for example, be warranted in holding the view that the property of having a kidney is identical to the property of having a heart. 'Has a kidney' and 'has a heart' are coextensive in our world, but the property of having a kidney is different from the property of having a heart.

model of other common identity claims such as 'water is H₂O', 'temperature is mean kinetic, molecular energy' and 'light is electromagnetic radiation'. The important features of these identity claims are that they are synthetic and necessary.

Brink here follows Kripke (Kripke 1980). The fact that an identity claim is synthetic does not rule out that it is also necessary. The pair of notions 'necessary' and 'contingent' are metaphysical notions and should neither be confused with the pair of notions 'a priori' and 'a posteriori', which are epistemological notions, nor with the pair of notions 'analytic' and 'synthetic' which are semantic notions. Natural kind terms each denote a natural kind that has an essence, and this essence is discovered through empirical enquiry. When an essence of a natural kind has been discovered, the true metaphysical nature of the kind has been revealed. This means that in any possible world in which instantiations of this natural kind have membership, these instantiations have this nature. The claim that 'water is H₂O' therefore expresses a necessary truth even though the claim is not analytically true.

Moving on to b), the question arises of what it means that moral properties are constituted by natural properties. Brink characterises the constitution relation like this: "If G actually composes or realises F, but F can be, or could have been, realised differently, then G constitutes, but is not identical with, F" (ibid.:157). To shed more light on the constitution relation, consider, for example, the relation between a bronze statue and the lump of bronze with a particular shape that takes up the same position in space as the statue. Is the bronze statue identical to the lump of bronze of that particular shape? The answer seems to be 'no' since the statue and the lump of bronze of that particular shape have different persistence conditions. Imagine that the statue is a full size depiction of Napoleon standing with his right hand in the air. Now, if you go and slightly bend one of the fingers on 'Napoleon's' right hand, then,

according to Brink's intuition, the statue will continue to exist whereas the lump of bronze of the particular shape will cease to exist. What we have after this act of vandalism is the same statue but a different lump of bronze. The lump of bronze is different because it is of a slightly different shape than it used to be. Because the statue can survive certain changes in the arrangement of bronze-molecules, it is the case that the statue could have been realised somewhat differently from the way it is actually realised.

The relation of constitution might be applied to the issue of what the relation is between natural and moral properties in the following way: one might say that though the latter are not identical to the former, they are nonetheless constituted by them. The reason for saying this would be the assumption that though moral properties are actually realised or composed of natural properties, moral properties could have been realised by properties that are not natural: say, non-natural or supernatural ones.⁸⁶ In other words, moral properties could have been realised differently from the way they are actually realised. Now, if one takes the view moral properties are natural properties in the sense that they are constituted by natural properties, then, according to Brink, one is debarred from taking the view that moral properties are identical to natural properties (*ibid.*:158;177). The argument is this:⁸⁷

- 1) If moral properties are identical to natural properties, then it is necessary that moral properties are identical to natural properties.
- 2) Moral properties, though actually constituted by natural properties could have been realised by a different type of properties.

⁸⁶On Brink's view, Moore construes goodness as a non-natural property because he construes it as a property that is ontologically *sui generis*.

⁸⁷Let us call this argument 'A'.

- 3) Therefore, it is not necessary that moral properties are identical to natural properties.
- 4) Therefore, it is not the case that moral properties are identical to natural properties.

The counterfactual assumption embedded in the second premise is, according to Brink, true.

"But it may seem that moral properties, even though actually realised by natural properties could have been realised by properties that are not natural, for instance supernatural properties" (ibid.:177).

In Brink's opinion, the above argument therefore provides the naturalist with a reason to resist the identification of moral and natural properties.⁸⁸ He offers yet another argument as to why the naturalist should resist this identification. The starting-point of this argument is the thesis that moral properties are functional properties that are multiply realisable. Brink's functionalism about moral properties is explicitly modelled on functionalism in the philosophy of mind:

"A plausible claim about a variety of property types and tokens (properties and property instances) is that they could have been realised in a variety of different ways. Functionalist theories of mind, for example, are based partly on this kind of claim about the one-many relationship between mental states and physical systems. [..]. A similar claim seems plausible about moral properties. For example, both the property of injustice and particular instances of injustice, in whatever social and economic conditions they are

⁸⁸ Brink says that identity implies constitution (ibid.:157). This is, however, not the case. Assume that being F is identical to being G. Assume also that identity implies constitution. On these assumptions, we are warranted in concluding that being F is constituted by being G. On Brink's view, the constitution relation is a contingent relation. If being F is constituted by being G, then it is possible that something should be F and not G. But then it cannot be the case that being F is identical to being G. The reason for this simply being that Brink construes identity as a necessary relation. If being F is identical to being G, then it is impossible that something should be F and not G.

actually realised, could have been realised by a variety of somewhat different configurations of social and economic properties and property instances" (ibid.:158).⁸⁹

How might the multiple realisability of moral properties provide a reason for resisting the identification of moral properties with natural properties? Brink writes:

"Moral properties could have been realised by an indefinite and perhaps infinite number of sets of natural properties. If we deny that identity is a relation that can hold between relata that are indefinitely or infinitely disjunctive - say, because we insist that identity holds only between genuine properties and we deny that disjunctive properties are genuine properties (cf. Armstrong 1978: II, 19-23) - then the multiple realisability of moral properties provides us with a reason for resisting the identification of moral properties and natural properties" (Brink 1989:158).⁹⁰

I do not think that this argument is entirely clear. Some of the confusion arises from the fact that Brink talks somewhat interchangeably about various kinds of relata. One might wonder what exactly the supposed problem is: is it that identity is a relation that cannot hold between relata that are 1) disjunctive, 2) indefinitely disjunctive or 3) infinitely disjunctive? There is, it seems to me, a problem with an identity claim that involves an indefinite disjunction as one of its relata. Consider the claim 'A is identical to B or to C or to.. '. Assume a standard truth-functional semantics for connectives under which the meaning of a connective is identical to the truth-table for a compound in which the connective is the main connective. Under such a semantics, we know the meaning of, say, 'A or B' just in case we know exactly which

⁸⁹Brink also advances functionalism in this passage: "The moral realist might claim that moral properties are functional properties. He might claim that what is essential to moral properties is the causal role which they play in the characteristic activities of human organisms" (Brink 1984:121).

⁹⁰Let us call this argument 'B'.

interpretations make the compound true and which makes it false. What is involved in the initial identity claim is an open sentence. Since the disjunction involved is open, we do not know exactly what interpretation makes it false. We therefore do not know the meaning of the disjunction, and as a result of this, we do not know the meaning of the identity claim in which it is involved.

Since an infinite disjunction is also an open sentence, the same argument can be used to show that an identity claim involving an infinite disjunction is problematical.

Is there also a problem with disjunctions per se? In other words, is there a problem with an identity claim that involves a (closed) disjunction as one of its relata? One might think so on the ground that whereas there is a predicate 'is A or B', there is no property of being A or being B. I think that there is good reason to reject the view that there are disjunctive properties. The argument is this: a plausible criterion for distinguishing between properties is a causal criterion. Property A is different from property B just in case an object instantiating A but not B, plays a role in the causal nexus that is different from the role played by an object that instantiates B but not A. This criterion explains the intuitive difference between, say, being a spoon and being a goldfish.

Imagine now that an object instantiates the property of being a spoon. Does this object also instantiate the property of being a spoon or a goldfish? On the causal criterion, the object instantiates this disjunctive property just in case its instantiation of it gives the object causal powers different from those it has in virtue of instantiating the property of being a spoon. This does not, however, seem to be the case. An object's (alleged) property of being a spoon or a goldfish does not give the object causal powers over and above those it has in virtue of being a spoon.

So, I think that there is reason to believe that there, in addition to a problem with indefinite and infinite disjunctions, is a problem with disjunctions per se. A disjunctive property is not a genuine property, and it seems plausible to say that identity is a relation that can only hold between genuine properties. Therefore, an identity relation cannot hold in cases where one (or both) of the relata is disjunctive.

Suppose that one denies the assumption that a disjunctive property is not a genuine property. One would then find the above argument unconvincing. Can Brink construct another argument that, without controversial claims about the status of disjunctive properties, takes us from the premise that M is multiple realisable to the conclusion that M should not be identified with any of the natural properties that might realise M on different occasions? Perhaps one is tempted to give an affirmative answer to this because of the availability of the following argument: Assume that M in the actual world is realised by N*. Since M is multiple realisable, there is, however, a possible world in which M is realised by N**. Suppose now that the identity theorist says that M is identical to N*, and to N**. Since identity is a transitive relation, this claim entails that N* is identical to N**. This is surely false, so anyone who thinks that M is multiple realisable should resist the identification of M with any of the natural properties that on different occasions might realise M.⁹¹ On the basis of arguments A and B, Brink thinks that the ethical naturalist should refer to

⁹¹ This argument is not conclusive. In my view, a plausible response to the idea that M is multiple realisable is to combine a type-type identity theory between M and some type of realiser property, say N*, with the idea of restricted identity (see Braddon-Mitchell & Jackson 1996:99). The property of being unjust (M) might be a functional property so that in different worlds, different combinations of social and economic properties (N*, N** etc.) realise M. This does not, however, seem to rule out the possibility of making true identity claims about the types of things that are injustice in different worlds. So, assume that M is realised by N* in the actual world, and that M is realised by N** in a different possible world. The identity theorists must then restrict himself to saying something like this: 'In the actual world, injustice is N*', and 'In possible world W₁, injustice is N**'. Analogously, what plays the pain role in humans might be C-fiber stimulation, and what plays the pain role in dolphins might be D-fiber stimulation. This is, however, compatible with true type-type identity claims of the following kind: 'Pain in humans is C-fiber stimulation', and 'Pain in dolphins is D-fiber stimulation'.

the 'is' of constitution when asked to explicate the defining claim of his position: that moral properties are natural properties.

"There is perhaps some reason, then, to construe ethical naturalism, or for that matter ethical supernaturalism, as a claim about the constitution rather than the identity of moral facts and properties. Moral facts and properties, so construed, are constituted, composed, or realised by organised combinations of natural and social scientific facts and properties. The former are, then, in a certain sense nothing over and above the latter" (Brink 1989:159).

3. Weak and Strong Supervenience

According to Brink, ethical naturalism involves the claim that moral properties supervene upon natural properties. Brink construes supervenience as

"a nomological or lawlike relation between, say, properties such that one property F (the supervening property) supervenes on another property, or configuration of properties G (the base property or properties), just in case it is a law that if something is G, it is F" (ibid.:160).

Brink makes a distinction between nomological relations that hold necessarily and nomological relations that hold contingently. He grounds a distinction between strong and weak supervenience in this distinction. On his view, ethical naturalism implies both strong and weak supervenience (ibid.:).

To understand Brink's notion of supervenience, we need to understand the notion of a law (or nomological relation) that is involved in his definition of supervenience. Take the suggestion that it is a law that if something is G, then it is F. If this suggestion is correct, then it is not only the case that as a matter of fact, everything G is F. Something stronger is the case: namely that it is impossible that

something should be G and not F. A modal aspect is essentially involved in a law or nomological relation. We can therefore give the following definition of a (specific) law: it is a law that if something is G, then it is F, iff it is necessary that if something is G, then it is F.

What is special about strong supervenience is that the base properties necessitate the supervening properties. What does this mean? Consider the claim that injustice (F) strongly supervenes on a set of social and economic properties (G*). G* necessitates F just in case it is true in all possible worlds and for all objects that if an object is G*, then it is F. Assuming that G* necessitates being F, it is the case that in the actual world, a social arrangement being G* is a sufficient condition for it being unjust. It is, however, also the case that in worlds that differ from the actual world in terms of properties causally connected to those in G*, being G* is still a sufficient condition for being F. What is involved in Brink's notion of strong supervenience can be put like this: α -properties (say, moral properties) strongly supervene upon β -properties (say, social and economic properties) iff:

- (i) For every F in α there is a property G* in β such that it is necessary that if something is G*, then it is F.⁹²

To cast further light on Brink's idea that moral properties strongly supervene upon natural properties, consider this passage: "Moral facts and properties strongly supervene upon natural facts and properties because some sets of natural properties necessitate certain sets of moral properties. Because he rejects a semantic test of properties, the ethical naturalist denies that this necessary relation represents either

⁹²Put formally: α -properties strongly supervene upon β -properties iff:

$$\forall F \in \alpha \exists G^* \in \beta [\Box (\forall x)(G^*x \rightarrow Fx)]$$

In this formula, the modal operators range over all metaphysically possible worlds, and every individual property in α and β is one that it is possible that an object should instantiate.

logical or conceptual necessity; instead, it represents metaphysical necessity or necessity a posteriori" (ibid.:175).

It is worth noticing that the idea that F strongly supervenes on G*, on Brink's construal of strong supervenience, is compatible with the idea that F is multiply realisable. The strong supervenience of F on G* leaves it possible that something should be F and not G*. Strong supervenience of F on G* only entails that being G* is a sufficient condition for being F in all possible worlds. It does not entail that being G* is necessary for being F (either in some possible world or in all possible worlds). I think that this brings out an important difference between the ethical naturalist who is an identity theorist and the ethical naturalist who is a constitution theorist.

Consider what an identity theorist (outside ethics) should say about the relation between being H₂O and being water. If asked to present his position in the strongest possible way, he should not say that it is metaphysically necessary for all objects, that if an object is a sample of H₂O, then it is a sample of water, because this would make his position consistent with it being metaphysically possible that something should be water and not H₂O. He does not, however, believe that this is metaphysically possible. For the identity theorist, being H₂O is not only a sufficient condition of being water: it is also a necessary condition. So, what the identity theorist should commit himself to, is the claim that it is metaphysically necessary for all objects that an object is a sample of water if and only if it is a sample of H₂O.

The ethical naturalist who is a constitution theorist should not commit himself to a universally quantified bi-conditional that holds with necessity. Where F and G* stand for the previously mentioned properties, the constitution theorist should not, that is, commit himself to the claim that it is metaphysically necessary for all

objects that an object is F if only if it is G*. Such commitment namely entails a commitment to the claim that it is metaphysically necessary for all objects that if an object is F, then it is G*. This is a commitment to the effect that F can only be realised by being G*, and this runs counter to the idea that being F is realisable in multiple ways.

According to Brink, it is often difficult to identify a strong supervenience base, because it is often difficult to be sure that a set of base properties includes all the properties required literally to necessitate a set of supervening properties. So, we often settle for identifying the weak supervenience base (ibid.:161). This is an epistemological, and not ontological point. The fact that it is often difficult to find out what the strong supervenience base is of a given moral property, does not mean that it does not have one.

Weak supervenience states a nomological relation that is not necessary. It holds in the actual world and in nearby possible worlds. Brink's notion of weak supervenience can formally be defined like this: α -properties weakly supervene on β -properties iff:

$$\forall F \in \alpha \exists G^* \in \beta [\Box (\forall x)(G^*x \rightarrow Fx)]^{93}$$

Here is what Brink says in order to explicate the distinction between strong and weak supervenience:

"Similarly, certain social and economic conditions may be sufficient in our world (and nearby possible worlds) to cause social injustice, even if there are possible societies sufficiently different from our own in certain respects that these social and economic conditions would not constitute injustice. Injustice,

⁹³ Importantly, in this formula, the modal operators do not range over the full set of metaphysically possible worlds. They only range over a subset of these worlds: namely the metaphysically possible worlds that are relevantly similar to the actual world. Again, every individual property in α and β is one that it is possible that an object should instantiate.

then, supervenes weakly, but not strongly, on these social and economic conditions" (ibid.:161).

Just to comment on this example: assume that in the actual world, a social arrangement's instantiation of G* is sufficient to make it F. Just in case there is one, or more, metaphysically possible worlds in which a social arrangement's instantiation of G* is sufficient to make it F, and that there is a metaphysically possible world in which a social arrangement's instantiation of G* is not sufficient to make it F, does F weakly supervene on G*. F's weak supervenience on G* is a nomological relation. It is not only the case that everything G* in our world is also F. It is impossible that something should be G* and not F where this means that there is no world in a certain subset of metaphysically possible worlds in which something is G* and not F.

An important question that arises in connection with Brink's construal of strong and weak supervenience is the question of how we find out which moral properties strongly, or weakly, supervene upon which natural properties. Brink's answer is here that this is something that is determined by engaging in first order moral inquiry. Different moral theories will give different answers to the question of what set of natural properties, a given moral property, say, injustice, supervenes upon.

"Which moral properties strongly supervene upon which natural properties is determined differently by different moral theories, just as, say, which economic properties strongly supervene upon which social and psychological properties is determined differently by different economic theories" (ibid:175).

4. Cornell Realism and Moral Supervenience

I have argued, in chapters two and three, that moral properties, at least weakly, supervene upon natural properties. I have, moreover, argued that any minimally convincing metaethical theory must be able to explain why this relation holds. What I mean by 'weak supervenience' is not, it is important to note, what Brink means by the term. In my use of the term, α -properties weakly supervene on β -properties just in case:

- (ii) Necessarily, for any property F in α and for any object x , if x is F , then there exists a property G in β such that x is G , and if any y is G , it is F .⁹⁴

What any minimally convincing metaethical theory must be able to explain is, in other words, why it is impossible that two objects within the same world should differ in terms of moral properties without differing in terms of natural properties. If Brink's view about the metaphysical status of moral properties is taken to represent the views of Cornell realism on the issue, then, by determining whether or not Brink can account for the supervenience relation, it can be determined whether or not Cornell realism can account for it. So, in order to determine whether or not Cornell realism is minimally convincing metaethical theory in the above mentioned respect, I now turn to consider whether or not Brink can account for the supervenience relation.

One important thing to focus on in connection with this issue is Brink's view that moral properties are constituted or realised by organised combinations of natural and social scientific properties. As Brink says, it is a consequence of the constitution view that moral properties are, in a certain sense, nothing over and above organised

⁹⁴ Put formally: α -properties weakly supervene on β -properties just in case:

$\Box \forall F \in \alpha \forall x [Fx \rightarrow \exists G \in \beta (Gx \ \& \ \forall y (Gy \rightarrow Fy))]$. This is Kim's definition of weak supervenience as discussed in chapter three.

combinations of natural and social scientific properties (ibid.:159). Now, if being F (where F is a particular moral property) is nothing over and above being constituted by a certain set of natural properties, then it is clear that two objects within a given world cannot differ in terms of being F without differing in terms of natural properties. On the constitution picture, a difference in terms of being F between two objects just is a difference in terms of their natural properties. It is therefore impossible that two objects within the same world should differ in terms of being F without differing in terms of natural properties. This is Brink's own explanation as to how his theory can account for the supervenience of moral properties on natural properties:

"One explanation of why moral properties supervene on natural ones is that they are constituted by, but are not identical with, complex configurations of natural properties. On this view, moral properties stand to natural properties much as a statue stands to the bronze out of which it is constituted. If so, there is one sense in which moral properties are nothing over and above the natural properties on which they supervene: the natural properties of a situation fix or determine its moral properties: the moral properties of a situation do not have to be added separately to the natural properties of the situation" (Brink 2001:157).⁹⁵

If nothing else follows from the constitution view of moral properties than that such properties are, in a certain sense, nothing over and above organised combinations of natural and social scientific properties, then Cornell realism, by advancing the above explanation, can give a convincing explanation of moral supervenience. However, it

⁹⁵This explanation of moral supervenience is also found in this passage: "Naturalists claim that moral properties supervene on natural properties *because* moral properties are constituted by natural properties" (Brink 1989:160).

is, arguably, not the case that the only thing that follows from the constitution view of moral properties is that such properties are, in a certain sense, nothing over and above natural properties. If this is right, then it cannot, as of yet, be concluded that Cornell realism can give a convincing explanation of moral supervenience. It might namely be the case that whatever it is that follows from the constitution view, in addition to moral properties being, in a certain sense, nothing over and above organised combinations of natural properties, is such that by accepting the constitution view, one incurs an explanatory burden that cannot be discharged by giving the above explanation. So, what needs to be addressed now is the question of what exactly follows from the constitution view, and whether or not this compromises the explanation of moral supervenience that Brink has offered.

5. The Real Explanatory Challenge

Brink's claim that the constitution view entails that, in a certain sense, moral properties are nothing over and above organised combinations of natural properties implies, though not strictly, that there is another sense in which moral properties are something over and above organised combinations of natural properties. That there is a sense in which this is the case is explicitly confirmed by Brink:

"In another sense, though, the moral properties are something over and above the natural properties on which they supervene. The same moral properties could have been realised by somewhat different configurations of natural properties; this modal difference implies that the properties are different properties, and grounds the constitution claim rather than the identity claim" (ibid.:157).

As previously mentioned, Brink's view on the metaphysical status of moral properties is modelled on a functionalist account of mental properties. Brink explicitly cites Putnam as someone who formulates a view about mental properties that is similar in structure to Brink's view about moral properties (Brink 1989:158). Here is what Putnam says, in the work referred to by Brink, about the mental property-type of being in pain:

"It would still be absurd to say, 'pain is stimulation of the copper fibers'. If we said that, then we would have to say that pain is something different in the case of machine I and in the case of machine II, if machine I had copper pain fibers and machine II had platinum pain fibers. Again, it seems clearer to say what we said before: that 'pain' *is* a state of the machine normally occasioned by damage to the machine's body and characterised by giving rise to inclinations' to...etc., and to eschew the formulation, 'Pain is synthetically identical with stimulation of the copper fibers..'" (Putnam 1967:420).

What Putnam is emphasising here, is that on the version of functionalism that he defends, the mental property-type of being in pain is identical to a certain internal state that is characterised in terms of the causal relations it enters into. This state may be realised in multiple ways, but it would be a mistake to think that being in pain, at the level of mental state-type, is identical to any first order property that realises this second order property of being in pain. The 'is' in "that 'pain' is a state of the machine.." is the 'is' of identity. So, Putnam is proposing the view that the mental state-type of being in pain is something over and above the first order properties that realise it on various occasions. This suggests that the version of functionalism that Putnam defends involves an identity claim at one level, namely that the mental property-type of, say, being in pain is identical to a certain type of internal state: a

state that might be realised in multiple ways. On the issue of whether or not functionalism in the philosophy of mind involves an identity-claim, one commentator says:

"Notice that functionalism in the philosophy of mind does involve an identity between mental properties and functional properties, the latter, of course, being realisable by a variety of lower level, more narrowly natural properties" (Timmons 1999:43).

Now, if Brink models his view on moral properties on, among others, Putnam's functionalism about mental properties, then it seems to me that Brink is committed to something like this: assume that empirical investigation has revealed that human uses of 'right' is causally regulated by the functional property of causing a maximisation of human well-being. This means that rightness is, where the 'is' is the 'is' of identity, this second order, functional property. An act is right if and only if it causes a maximisation of human well-being. An act's property of being right is realised by other, first order properties of the act, just as well as, on Putnam's view, a creature's property of being in pain is realised by certain other properties of the creature: e.g. C-fiber stimulation. Rightness is not identical to any first order, realiser property, and an act's property of being right is therefore, in an important and straight forward sense, something over and above its other, first order properties that realise it.

This account of the status of moral properties fits well with the semantic account, discussed in the previous chapter, that Cornell realism gives of moral predicates. What 'right' rigidly designates is the second order, functional property of causing a maximisation of human well-being. 'Right' does not, it is important to note, rigidly designate first order, realiser properties.

"According to *synthetic moral functionalism*, as we will call this position, moral terms are construed as purporting to pick out certain functional properties whose essence is revealed by the generalisations of a synthetic moral theory" (Horgan & Timmons 1992a:242).

The fact that this account of the metaphysical status of moral properties entails that, in one sense, moral properties are something over and above the first order, natural properties that realise them, makes it the case that Brink's previous explanation of moral supervenience is not one that relieves him from problems in relation to moral supervenience. The previous explanation of moral supervenience only works on the assumption that moral properties are nothing over and above organised combinations of natural properties. It has now emerged that, in one important sense, this assumption is false. The explanatory challenge that Brink faces should initially be formulated like this: how is it possible for someone who thinks that moral properties are second order, functional properties that are ontologically different from the first order, natural properties that realise them, to explain why it is impossible that two objects in the same world should differ in terms of moral properties without differing in terms of natural properties?

6. *Brink's Response*

One possible answer to this explanatory challenge is to say that (i) is true. The truth of (i) entails the truth of (ii). Put differently, it is a consequence of strong supervenience, in Brink's sense of strong supervenience, that there is no possible world in which two objects differ in their moral properties without differing in terms of their natural properties. Let me try to make this clear:

Assume that there is in the actual world an object that instantiates moral property F (say, injustice). Since moral properties on Brink's view are realised by first order, natural properties, the object in question also instantiates such a set of properties. Let us assume that F on this particular occasion is realised by property G*. Let us also assume that G* is a strong supervenience base for F. F may have more than one strong supervenience base. Intuitively, it may be the case that in all possible worlds, being G* is sufficient for being F. This does not, however, rule out that there is a different configuration of social and economic properties (G**) that is also sufficient, in all possible worlds, to cause F.

Now, if G* is a strong supervenience base for F, then there is no possible world in which something is G* and not F. This explains why moral supervenience, in my sense of moral supervenience, holds for the particular moral property F. If there is no possible world in which something is G* and not F, then of course there is no possible world in which something is G* and F, and something else is G* and not F. A similar argument can be given as to why moral supervenience holds for any other moral property.

Consider this definition of strong moral supervenience: α -properties (moral properties) strongly supervene on β -properties (natural properties) just in case:

- (iii) Necessarily, for any property F in α and for any object x, if x is F, then there exists a property G in β such that x is G, and necessarily if any y is G, it is F.⁹⁶

It is important to note that Brink is committed to the view that moral properties strongly supervene upon natural properties, in this sense of strong moral

⁹⁶ Put formally: α -properties strongly supervene on β -properties just in case:

$\Box \forall F \in \alpha \forall x [Fx \rightarrow \exists G \in \beta (Gx \ \& \ \Box \forall y (Gy \rightarrow Fy))]$. This is Kim's definition of strong supervenience as discussed in chapter three.

supervenience. On the assumption that there is something in the actual world that is F and whose strong supervenience base is G*, Brink is committed to the view that there is no object in either the actual world or in any other world that is G* and not F. But again, this is easily explained by the fact that there is no possible world in which something is G* and not F.

In this section, I have so far left open the question of what the force is of the modal operators involved in the various supervenience claims. It is now time to be more precise in this respect. As noted, Brink is of the opinion that (i) holds with metaphysical necessity. So Brink can explain why there is no metaphysically possible world in which two objects are identical in terms of first order, realiser (natural) properties, and yet differ in terms of moral properties. This is not insignificant, but it is not enough to relinquish Brink from problems in relation to moral supervenience. As I have tried to argue in chapters two and three, moral supervenience holds with conceptual necessity. It is a conceptual constraint on any user of moral language, and no explanation of this conceptual constraint flows naturally from the claim that (i) holds with metaphysical necessity. Metaphysical necessity does not entail conceptual necessity, so the fact that Brink can explain why there is no metaphysically possible world in which two objects are identical in terms of first order, realiser (natural) properties, and yet differ in terms of moral properties does not entail that he can explain why there is no conceptually possible world in which this state of affairs obtain.

If Brink took it to be the case that (i) holds with conceptual necessity, then he could explain moral supervenience. However, to think that (i) holds with conceptual necessity is equivalent to an abandonment of synthetic naturalism in favour of

analytic naturalism. Brink cannot therefore explain moral supervenience by changing his view about the modal force of (i).

Is it possible to construct an argument that both allows Brink to explain moral supervenience as a conceptual constraint, and allows him to retain his claim that (i) holds with metaphysical necessity? Blackburn does not think so:

"This may indeed be the kind of thing that a certain kind of 'realist' (Cornell realism) might say, except that they would want to avoid an identity of 'conceptual' content. But then the idea that it is satisfactory comes from focusing exclusively on reference to properties. Generally speaking, identity of properties does not explain the conceptual requirement that supervenience imposes in the case of ethics. For example, people say that the property of being water is identical with the property of being H₂O. But this does not make it a priori that being water supervenes on being H₂O, or indeed on any chemical composition [...]. It is the analyticity of the supervenience requirement, in ethics, that is the awkward problem for realism. Reference to properties does not explain it. If the realist claims identity of conceptual content, between 'is good' and 'creates happiness' he escapes the problem, but the open question argument strikes" (Blackburn 1998:316).

In my opinion, Blackburn is too quick in his dismissal of Cornell realism.⁹⁷ Here is what Brink should do in order to fend off Blackburn's criticism. He should start by insisting on a separation of issues. He should then draw attention to the fact that at the level of metaphysical necessity, he can explain everything that needs to be explained in relation to moral supervenience. The fact that moral supervenience is a

⁹⁷In my mind, he is also too quick in his dismissal of a theory (analytic ethical naturalism) that claims identity of content between moral and natural predicates. Such a theory might be mistaken, but as I tried to argue in chapter eight, there is no reason to think that the open question argument can be used to strike it out.

conceptual constraint on any user of moral language is, however, a separate issue that needs an explanation that is separate from the one given as to why there is no metaphysically possible world in which two objects are identical in terms of first order, realiser (natural) properties, and yet differ in terms of moral properties.

This separate explanation involves considerations about what semantic role is of moral language. Put shortly, there is no reason why Brink should not be allowed to explain the analyticity of moral supervenience by copying, to some extent, the explanation that Blackburn (and Hare) gives of the phenomenon. Brink gives a descriptivist account of moral discourse. On this account, moral language is primarily used to give a description of an aspect of reality. What lies at the heart of an advanced debate between proponents of a non-descriptive analysis of moral discourse and proponents of a descriptive analysis of moral discourse is not an issue of whether or not moral discourse is entirely non-descriptive or entirely descriptive. As Blackburn himself reminds us, an advanced non-descriptivist (expressivist) does not commit a 'speech act fallacy' that consists in thinking that moral predicates have an exclusively non-descriptive meaning (Blackburn 1984:170). At least since Hare, it has been part of non-descriptivist orthodoxy to allow that moral predicates have an element of descriptive meaning. To apply a moral predicate to an object is not only to prescribe or recommend it, or otherwise express something with non-propositional content: the defining claim of non-descriptivism is only that the primary meaning of moral predicates is non-descriptive.

Formulated as a negative thesis, descriptivism is the view that the non-descriptivist view of moral language is mistaken. Formulated as a positive thesis, descriptivism is the view that moral discourse is primarily descriptive. A descriptivist is therefore not committed to the view that moral language does not

posses an element of non-descriptive meaning. This means that Brink qua descriptivist is allowed to say that fundamental moral terms such as 'good', 'right', 'unjust' etc., have a partly non-descriptive meaning. In other words, there is nothing in descriptivism as such that debar Brink from saying that part of what we do when we apply moral predicates to objects is something else than giving a description of these objects. What might this 'something else' be? Here it seems natural to say that we express recommendations, teach standards, and try to guide practical decision-making. Non-descriptivist are, it seems to me, forced to allow the descriptivist this hybrid view of moral language. Otherwise, the idea that they can allow themselves a hybrid view, seems equivalent to be introducing double standards.

Now, if Brink is entitled to say that we use moral language, at least partly, to teach standards and guide practical decision-making, then he can, it seems to me, explain moral supervenience as a conceptual constraint. He can do this by simply copying the expressivist's explanation of moral supervenience. He can argue that moral language, at least partly, serves the role of being a vehicle by which we teach standards and guide practical decision-making. Our use of moral predicates could not serve this role if we did not respect supervenience in our practice of applying these predicates. It is therefore not surprising that it is part of the content of moral concepts (and part of the meaning of the predicates that express these concepts) that they are to be applied uniformly across descriptive similarities.

Where the modal operator denotes conceptual necessity, Brink can therefore accept and, importantly, explain the truth of the claim that α -properties (moral properties) weakly supervene on β -properties (first order, realiser properties) in the sense that: $\Box \forall F \in \alpha \forall x [F x \rightarrow \exists G \in \beta (G x \ \& \ \forall y (G y \rightarrow F y))]$

7. Final Remarks

It seems to me that the considerations of the previous section support the conclusion that Brink (and therefore Cornell realism) does not have a problem on the issue of moral supervenience. The suggestion that Cornell realism fares badly on this issue arises from a failure to appreciate two things: firstly that the explanation that Cornell realism gives as to why it is metaphysically impossible that two objects within the same world should differ in terms of moral properties without differing in terms of first order, realiser properties, is not the only explanation that Cornell realism can, and should, give on the general issue of moral supervenience.

Secondly, that the descriptivist analysis of moral discourse that Cornell realism favours, is compatible with the view that moral concepts has, as part of their meaning, a certain non-descriptive content. When these two things are fully appreciated, it should be clear that a mistake is made if one, as Blackburn, supposes that the analyticity of the supervenience requirement in ethics poses a problem for Cornell realism.

Chapter 11

Conclusion

1. *Introductory*

This is the final chapter of the thesis. Section two contains a short summary of the preceding chapters and their main conclusions. In section three, I outline and discuss various ways in which these conclusions might be resisted. The last section is concerned with the question of what the connection is between metaethics and normative ethics. I address this in order to place the issues discussed throughout this thesis in their broader philosophical context.

2. *Summary and Main Conclusions*

This thesis has been concerned with metaethics. In the introduction, a distinction was made between external and internal accommodation projects for moral discourse. It was then argued that the external accommodation project should be guided by acceptance of methodological naturalism. After having made clear what methodological naturalism is and given a rationale for accepting it, a prolonged discussion was undertaken of two metaethical theories whose ontological and epistemological commitments are compatible with those of methodological naturalism: namely expressivism, as developed in the works of Simon Blackburn, and Cornell realism, as presented by Richard Boyd and David Brink.

In the course of the thesis these two theories have been subjected to an extended comparative evaluation, and an answer has been sought to the question of

which of the two should be favoured. My answer to this question is that Cornell realism is rationally preferable to expressivism. This is the main conclusion of the thesis. It was arrived at by looking at how well the two theories, respectively, explain various deeply embedded features of moral discourse and practice. Explaining such features is what the internal accommodation project for moral discourse consists in.

The assertoric surface-form of moral discourse and the supervenience of moral predicates on natural predicates are two features of moral discourse that have been the object of special attention in the thesis. It has been argued that whereas expressivism has no problems on the issue of moral supervenience, it still has no convincing answer to the challenge posed by the Frege-Geach point. In the current state of philosophical development, therefore, expressivism is not a metaethical theory to which a methodological naturalist should subscribe.

After the examination and assessment of expressivism, attention was directed at realist alternatives to it. Analytical ethical naturalism was first discussed, and here there was a special focus on the question of whether or not 'the open question argument' can be used to refute this position. Arguments were presented for a negative reply to this question. Cornell realism then became the centre of attention. First, an inquiry was made into the semantic views commonly associated with this position. An influential and much discussed objection to these views was considered in detail, and it was concluded that this objection cannot be used to show the falsity of the semantic component of Cornell realism. It was argued that what makes this conclusion interesting is the fact that the considerations upon which it is based, show that if the semantic component of Cornell realism is false, then it is so for a reason different from the one most commonly cited in the literature.

Some of the key metaphysical commitments of Cornell realism were then outlined, and a discussion was undertaken of the explanation that Cornell realism gives of moral supervenience. An issue that received special attention in this discussion was that of whether the particular modal force with which moral supervenience holds, poses a problem for a naturalistic theory that denies the existence of an analytic entailment relation between an object's instantiation of certain natural properties and its instantiation of certain moral properties. The conclusion of the discussion was that Cornell realism has no problems on the issue of moral supervenience.

In a comparative evaluation involving the major issues I have chosen for discussion, Cornell realism therefore fares better than expressivism. Expressivism and Cornell realism do equally well on the issue of moral supervenience. But whereas expressivism is vulnerable to a particular argument from the philosophy of language (the Frege-Geach point), Cornell realism can fend off the criticism that most persistently has been directed at it from this area of philosophy.

3. Resisting the Conclusions and a Short Glance Ahead

What are the prospects for resisting the conclusions of the thesis? One line of thought consists in saying that the judgement that Cornell realism should be preferred over expressivism, is of limited theoretical interest since it is false to assume that methodological naturalism should guide the external accommodation project for moral discourse. According to this objection, methodological naturalism imposes unacceptably tight restrictions on what count as plausible epistemological and ontological commitments. If one abandons the requirement that a metaethical theory should be compatible with methodological naturalism, then a number of other

metaethical theories emerge as alternatives to the two theories discussed in this thesis. It is therefore premature to suggest that Cornell realism should be favoured.

This line of argument features in the influential metaethical works of John McDowell.⁹⁸ It is also found in the work of writers such as Sabina Lovibond and Mark Platts.⁹⁹ This is not the occasion to engage in a discussion of the works of these writers. I mention them, however, in order to draw attention to the fact that if one goes against my recommendation and rejects methodological naturalism in favour of what might be labelled 'expansive naturalism', then there are writers on the contemporary metaethical scene in whose works one might find inspiration and intellectual affinity. Interesting as their work may be, however, it carries the significant burden of either having to reject naturalism in general or else of showing that its understanding of an expanded version of this meets the metaphysical and epistemological requirements of a credible metaethical theory. This remains the major challenge to non- or expansive naturalistic moral realism.

Another way in which one might try to resist the conclusions of the thesis is to say that whereas methodological naturalism should be accepted, the issues I have chosen for discussion for a comparative evaluation of expressivism and Cornell realism are not the important ones, or are ones that ought to be supplemented with various other issues. One might then go on to say that if the focus of attention is shifted from the set of discussed issues to others, then it is not at all obvious that the conclusion of the discussion stays the same: i.e. it is not at all obvious that Cornell realism has the upper hand in a comparative evaluation with expressivism. A defender of expressivism could, for example, say that I have ignored the important issue for moral psychology of what the connection is between having a moral

⁹⁸ See, for example, (McDowell 1979; 1985; 1995).

⁹⁹ See (Lovibond 1983) and (Platts 1979).

conviction and being motivated to act in accordance with it. Moreover, expressivism, supposedly, fares well on this issue and Cornell realism may have problems in relation to it. So, were this issue to be brought into the overall equation, expressivism would, arguably, come out as a stronger position than it appears in my discussion of it.

Let me comment on this line of thought. The overall structure of the thesis is given by the following question: if we accept methodological naturalism, have a choice between expressivism and Cornell realism and restrict ourselves to a discussion of issues x, y and z, which of the two theories should then be favoured? Evidently, this structure implies that if a change were made from a discussion of issues x, y and z, to an examination of a different set of issues, then it is possible that the overall conclusion would also change. However, the issues I have selected for discussion are not chosen arbitrarily or idiosyncratically. They connect to deeply embedded features of moral discourse that any minimally convincing metaethical theory must be able to account for. So, while the thesis does not constitute a fully comprehensive comparative evaluation involving all conceivably relevant features of moral discourse, its selectivity is considered, and I believe warranted by the importance, complexity and difficulty of the issues chosen for analysis.

Considerations about the set of issues selected for discussion suggest how one might proceed next in the general search for a sound metaethics. A natural progression from the thesis would indeed consist in looking at other deeply embedded features of moral discourse, and seeing how well the two theories, respectively, can explain these. A different line of inquiry, however, would be to direct attention at other metaethical theories that are compatible with methodological naturalism, but which have not been discussed at any length in the present study. I

mentioned some of these theories in section two of chapter one. In light of what was concluded in chapter eight following the discussion of the 'open question argument', analytic naturalism is likely to be of particular interest. In a series of articles in the 1950's and '60's, Philippa Foot defended a version of analytic naturalism.¹⁰⁰ She argued that there was a conceptual link between an object's possession of certain natural properties and its possession of certain moral properties. It would simply constitute a lack of competence with the involved predicates, if one failed to see that acts with certain natural features were rude or kind. In quite recent times, analytic ethical naturalism has been the subject of renewed interest among philosophers dissatisfied with expressivism and non-reductive realism. A significant reason for this is no doubt that the limitations of the 'open question argument' has become more and more apparent. As mentioned in chapter eight, Michael Smith and Frank Jackson are examples of contemporary defenders of analytic naturalism. There are at least two reasons as to why their theories are interesting. First, as versions of descriptivism, they ought to have no problems in accounting for the assertoric surface-form of moral discourse. Second, since the relation between being G (where G is some particular natural property) and being F (where F is some moral property) is held to be analytic, such theories seem well-placed to account for the special modal force with which moral supervenience holds. Whether or not these presumptions hold true, and whether or not Smith's and Jackson's theories can explain other deeply embedded features of moral discourse are, however, complicated questions that would need to be addressed in a separate and lengthy piece of work.

¹⁰⁰(Foot 1958a; 1958b; 1961)

4. *Metaethics and Normative Ethics*

Let me end by saying something about the connection between metaethics and normative ethics (moral theory and applied ethics). A reader of this thesis may note that it only discusses issues of an abstract theoretical nature, and that no effort has been made to connect these issues to issues in normative ethics. In light of certain developments within philosophical ethics in recent years, this feature of the thesis might appear puzzling, perhaps even worrying. The developments I have in mind are well summarised in this passage from the introduction to a contemporary textbook in ethical theory:

"Present-day philosophers, unlike the ancients, still employ these categories [Metaethics; Normative Ethics; Practical Ethics] to roughly distinguish the types of inquiry in which they engage. But, unlike their mid-century predecessors, they reject the idea that philosophy simply concerns "theoretical statements....about theoretical statements." They are therefore disinclined to think of these ethical categories as separate - as three wholly independent inquiries. For instance, Stephen Darwall rejects any clear separation between metaethics and normative ethics (1998:12), while Shelly Kagan not only eschews the distinction between metaethics and normative ethics (1998:7), he also renounces any firm distinction between normative and practical ethics" (Darwall 2000:2).

To the extent that this is a correct depiction of the present-day orthodoxy on the issue of what the connection is between metaethics and normative ethics, I am not part of this orthodoxy. In particular, I am not disinclined to think of these two ethical categories as separate: and I think that a relatively strong case can be made for the suggestion that metaethics is independent of normative ethics.

To say that A is independent of B is vague. It immediately raises the question: 'in what sense is A independent of B?'. So, if my suggestion that metaethics is independent of normative ethics is to be of interest, it needs to be qualified. Put more precisely, what I mean is that metaethics is independent of normative ethics in the sense that when one considers well-known and commonly held metaethical theories, then it is most often the case that their adoption has no implications for what normative views one can consistently hold. I am not claiming to be able to show that it is impossible that there should be metaethical theories that have normative implications. Candidates for such theories are, however, rare and there are questions as to whether they do not so much bridge the metaethical and the ethical as reject (or overlook) metaethics conceived of in metaphysical or semantic terms.

There is only one way of testing the plausibility of my suggestion and that is to consider a number of well-known and commonly held metaethical theories and see whether they have normative implications. If it can be shown that a number of these theories are such that the adoption of them has no implications for what normative views one can consistently hold, then this strengthens the credibility of the suggestion that metaethics is independent of normative ethics.

Consider first Blackburn's quasi-realist position. This is a metaethical stance in so far as it involves certain semantic, ontological and epistemological assumptions about the nature of moral discourse. This position does not entail any first-order moral commitments. Neither does it entail any particular moral theory. Addressing the issue of that the connection is between projectivism and consequentialism, or any other moral theory, Blackburn says:

"It should be noted from the outset that there is no essential connection between projectivism and a consequentialist view in ethics" (Blackburn 1985b:164).

So, if one focuses solely on quasi-realism (or, alternatively, on Ayer's emotivism or Hare's early formulation of prescriptivism), then it would, I believe, be correct to say that metaethics is independent of normative ethics.

Consider next a divine command theory according to which moral goodness is a property that an object instantiates in virtue of being commanded or otherwise endorsed by a Divinity. A divine command theory is a metaethical theory because it involves certain semantic, ontological and epistemological assumptions about the nature of moral discourse. One might think that this metaethical theory sets constraints on what normative views one can consistently hold. Assume that the Divinity does not command actions that maximise the overall amount of experienced pleasure. On this assumption, one cannot both be a divine command theorist and, in normative ethics, be a utilitarian who thinks that what is characteristic of right acts is that they maximise the overall amount of experienced pleasure. So, perhaps one can cast severe doubt on the plausibility of my suggestion about the independence of metaethics by making reference to a divine command theory.

Such a doubt would, however, be premature. A distinction should be made between a divine command theory as an account of the status of moral values and presumptions, and claims about what the content is of what is deemed good or what acts are commanded or prohibited. Claims of the latter kind belong to normative ethics, and not metaethics. A divine command theory does not in itself set constraints on what normative views one can consistently hold. Here it is useful to keep in mind the distinction between the question of 'what it is to be good', and the question of

'what is good'. A divine command theory is a theory that answers the former question, and not the latter. For any defender of a divine command theory there is a distinct philosophical task that consists in explaining and defending the various semantic, ontological and epistemological assumptions upon which his theory rests. A task of this nature is something quite different from the task of finding out what the content is of the Divinity's commands.

Finally, consider an ideal observer theory according to which what it means to say that something is good, is that it is what an ideal observer would favour. This is a metaethical theory because it involves certain semantic, ontological and epistemological assumptions about the nature of moral discourse. Consider, for example, the fact that such a theory involves a commitment to the idea that moral utterances are truth-apt. It also involves a commitment to a descriptivist analysis of moral utterances. So, on these two issues, an ideal observer theory involves commitments that are similar to those associated with, say, Moorean realism and Cornell realism. It differs, however, from the latter two theories when it comes to the question of what the truth-maker is of moral utterances. On this issue, an ideal observer theory appeals to the prescriptions of an epistemologically privileged agent. This is an appeal that neither Moorean realism nor Cornell realism make.

Does an ideal observer theory in itself set constraints on what normative views one can consistently hold? I do not think so. Consider again the distinction between the question of 'what it is to be good', and the question of 'what is good'. An ideal observer theory is a theory that answers the former question, and not the latter. The view that what it is to be good is to be favoured by an ideal observer is compatible with all kinds of views about what such an observer would actually favour, and the task of finding out what such an observer would favour is different

from the one that consists in explaining and defending the semantic, ontological and epistemological assumptions upon which a ideal observer theory rests.

Having addressed the issue of whether or not metaethics is independent of normative ethics and having offered an argument as to why it is independent, it is possible to move on to the question of why it is worthwhile to do metaethics. Why is it valuable to engage in this kind of inquiry? Imagine for a moment that one was convinced that metaethics is not independent of normative ethics. A particular answer to the question presently under consideration would then seem natural. This would be the answer that it is worthwhile to do metaethics because the adoption of a metaethical theory very often has implications for what normative views one can hold. The assumption of this answer is that the question of what normative views to hold is of great importance to us, and it is therefore important and interesting to know about whatever has implications for us in this area of thought.

Since I have argued for the suggestion that metaethics is independent of normative ethics, the above answer is not available to me. Moreover, the question as to why it is worthwhile to do metaethics might be seen to exert a special kind of pressure on someone who holds the view that metaethics is independent of normative ethics. The underlying assumption here is that if one rejects the idea that it is often the case that the adoption of a metaethical theory has implications for what normative views one can hold, then it is difficult to see what the justification could be for an interest in metaethics.

My answer to the question of why it is worthwhile to do metaethics goes via more general considerations. Consider first the nature of philosophical inquiry. One of its chief aims is to produce a coherent and systematic explanation of phenomena that we only partially understand, and/or have problematic or conflicting

explanations of. On this understanding, metaethical inquiry is a typical form of philosophical inquiry. What we aim at when we do metaethics is to provide a coherent and systematic explanation of that phenomenon that can be labelled 'moral discourse'. It is difficult to give a clear and concise criterion by which one can decide whether or not some statement or thought falls within the boundaries of moral discourse. It would, however, be mistaken to think that just because it is difficult to give such a criterion it follows that there is no intelligible distinction to be made between moral discourse and other areas of discourse. Moral discourse, arguably, has features that demarcate it from other areas of discourse. As I have argued several times over the course of this thesis, moral predicates supervene on natural properties with conceptual necessity. In this sense, moral discourse is different from, say, discourse about the mental. No conceptual mistake is made if one suggests that two persons within a world can differ in terms of some mental feature without differing in terms of some physical property.

Moral discourse is also different from, say, colour discourse. It is a well recognised fact that people are, other things being equal, motivated to act in accordance with their moral judgements. Michael Smith brands this characteristic feature of moral judgements, the 'practicality of moral judgement', and summarises it as follows:

"Moral judgements seem to be, or imply, opinions about the reasons we have for behaving in certain ways, and, other things being equal, having such opinions is a matter of finding ourselves with a corresponding motivation to act" (Smith 1994:7).

The 'practicality of moral judgements' gives rise to various views about what the exact nature is of the relation between giving a favourable moral judgement of

something and being motivated to do, or endorse it. Internalists in this area hold that the relation is necessary (See, for example, Blackburn 1984), whereas externalists hold that it is contingent (See, for example, Railton 1989).

A similar debate does not arise in connection with colour discourse. It is not a characteristic feature of colour discourse that people are, other things being equal, motivated to act in accordance with their colour judgements. For a philosopher engaged in a second-order inquiry about colour discourse, there is no issue of the 'practicality of colour judgements' that needs to be addressed.

If it is correct to suggest that there is such a thing as moral discourse, and that metaethical inquiry aims at producing a coherent and systematic explanation of this phenomenon, then it is possible to provide an answer to the question of why it is worthwhile to engage in metaethical inquiry. The answer I favour rests on the assumption that just as it is philosophically desirable in itself to have a systematic and coherent account of what goes on when we engage in areas of discourse such as mental, modal and colour discourse, so it is desirable in itself to have a similar account of what goes on when we engage in moral discourse. This assumption, together with the assumption that metaethical inquiry aims at producing a coherent and systematic account of moral discourse, implies that metaethical inquiry is worthwhile in its own right. Such is also my conviction and such is what has motivated this inquiry into what account of the nature and status of moral discourse best commends itself as explaining central and important features of that discourse.

Bibliography

Armstrong, D.A. 1980: "Naturalism, Materialism, and First Philosophy". In his *The Nature of Mind*. St Lucia, Queensland: University of Queensland Press, 149-165.

Arrington, R. L. 1989: *Rationalism, Realism, and Relativism*. Ithaca: Cornell University Press.

Ayer, A. [1936] 1946: *Language, Truth, and Logic*, 2nd edn. London: Gollancz.

Baldwin, T. 1993: Editor's introduction to Moore 1993.

Belnap, N. D. 1962: "Tonk, Plonk and Plink". *Analysis* 22, 130-34.

Blackburn, S. 1973: "Moral Realism". Casey, J. (ed.), *Morality and Moral Reasoning*. London Methuen. Reprinted in Blackburn 1993a. All page references to this text.

Blackburn, S. 1984: *Spreading the Word*. Oxford: Oxford University Press.

Blackburn, S. 1985a: "Supervenience Revisited". Hacking, I. (ed.), *Exercises in Analysis*. Cambridge: Cambridge University Press. Reprinted in Blackburn 1993a. All page references to this text.

Blackburn, S. 1985b: "Errors and the Phenomenology of Value". Honderich, T. (ed.), *Morality and Objectivity: Essays in Honour of J. L. Mackie*. London: Routledge & Kegan Paul.

Blackburn, S. 1988: "Attitudes and Contents". *Ethics*, Vol. 98, 501-517. Reprinted in Blackburn 1993a. All page references to this text.

Blackburn, S. 1992: "Gibbard on Normative Logic". *Philosophy and Phenomenological Research* 52 (4), 947-52.

Blackburn, S. 1993a: *Essays in Quasi-Realism*. Oxford: Oxford University Press.

Blackburn, S. 1993b: "Realism, Quasi or Queasy?". Haldane, J. & Wright, C. (eds.), *Reality, Representation, and Projection*. Oxford: Oxford University Press, 365-83.

Blackburn, S. 1993c: "Addendum" (to "Supervenience Revisited"), in Blackburn 1993a.

Blackburn, S. 1998: *Ruling Passions*. Oxford: Clarendon Press.

Blackburn, S. 1999: "Is objective Moral Justification Possible on a Quasi-realist Foundation?". *Inquiry* 42, 213-28.

Blackburn, S. 2002: "Replies". *Philosophy and Phenomenological Research* 65, 164-76.

Bloomfield, P. 2003: "The Rules of "Goodness": An Essay on Moral Semantics". *American Philosophical Quarterly*. Vol. 40, Number 3. July issue.

Boyd, R. 1988: "How to Be a Moral Realist". Sayre-McCord, G. (ed.), *Essays on Moral Realism*. Ithaca, NY: Cornell University Press, 181-228.

Boyd, R. 2003: "Finite Beings, Finite Goods: The Semantics, Metaphysics and Ethics of Naturalist Consequentialism, part 1". *Philosophy and Phenomenological Research* 67, 24-47.

Braddon-Mitchell, D. & Jackson, F. 1996: *Philosophy of Mind and Cognition*. Oxford: Blackwell Publishers.

Brighouse, M. 1990: "Blackburn's Projectivism - An Objection". *Philosophical Studies* 59, 225-35.

Brink, D. 1984: "Moral Realism and Sceptical Arguments From Disagreement and Queerness". *Australasian Journal of Philosophy* 62, 111-25.

Brink, D. 1989: *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press.

Brink, D. 2001: "Realism, Naturalism, and Moral Semantics". *Social Philosophy and Policy*, Vol. 18.

Copp, D. 2000: "Milk, Honey, and the Good Life on Moral Twin Earth". *Synthese* 124, 113-37.

Crisp, R. 1996: "Naturalism and Non-Naturalism in Ethics". Lovibond, S. & Williams, S. G. (eds.), *Essays for David Wiggins*. London: Routledge Press.

Darwall, S. 1998: *Philosophical Ethics*. Oxford: Westview Press.

Darwall, S, Gibbard, A, and Railton, P. 1992: "Toward *Fin de Siécle* Ethics: Some Trends". *Philosophical Review* 101, 115-89.

Dreier, J. 1992: "The Supervenience Argument Against Moral Realism". *The Southern Journal of Philosophy*, vol. XXX, no. 3.

Foot, P. 1958a: "Moral Arguments". *Proceedings of the Aristotelian Society* 59, 502-513.

Foot, P. 1958b: "Moral Beliefs". *Proceedings of the Aristotelian Society* 59, 83-104.

Foot, P. 1961: "Goodness and Choice". *Proceedings of the Aristotelian Society* , Supplementary vol. 25.

Gampel, E. 1997: "Ethics, Reference, and Natural Kinds". *Philosophical Papers* 26, 147-63.

Geach, P. 1960: "Ascriptivism". *Philosophical Review* 69, 221-5.

Geach, P. 1965: "Assertion". *Philosophical Review* 74, 449-65.

Gibbard, A. 1990: *Wise Choices, Apt Feelings*. Cambridge, MA: Harvard University Press.

Grayling, A. C. 1990: *An Introduction to Philosophical Logic*. London: Duckworth.

Hale, B. 1986: "The Compleat Projectivist". *Philosophical Quarterly* 36, 65-84.

Hale, B. 1993a. "Can There Be a Logic of Attitudes". Haldane, J. & Wright, C. (eds.), *Reality, Representation, and Projection*. Oxford: Oxford University Press, 337-63.

Hale, B. 1993b: "Postscript". Haldane, J. & Wright, C. (eds.), *Reality, Representation, and Projection*. Oxford: Oxford University Press, 385-8.

Hale, B. 2002: "Can Arboreal Knotwork Help Blackburn Out of Frege's Abyss?". *Philosophy and Phenomenological Research*, 65, 144-9.

Hare, R. M. 1952: *The Language of Morals*. Oxford: Oxford University Press.

Hare, R. M. 1984: "Supervenience". *Proceedings of the Aristotelian Society* 58, Supplement, 1-16.

Harman, G. 1977: *The Nature of Morality*. Oxford: Oxford University Press.

Heil, J. 1998: "Supervenience Deconstructed". *European Journal of Philosophy*, Vol. 6, number 2.

Hintikka, J. 1969: "Deontic Logic and Its Philosophical Morals". In his *Models for Modalities*. Dordrecht: Reidel.

Hooker, B. (ed.) 1996: *Truth in Ethics*. Oxford: Blackwell.

Horgan, T. & Timmons, M. 1991: "New Wave Moral Realism Meets Moral Twin Earth". *Journal of Philosophical Research* 16, 447-65.

Horgan, T. & Timmons, M. 1992a: "Troubles On Moral Twin Earth: Moral Queerness Revived". *Synthese* 92, 221-60.

Horgan, T. & Timmons, M. 1992b: "Troubles For New Wave Moral Semantics: the 'Open-Question Argument' Revived". *Philosophical Papers* 21, 153-75.

Jackson, F. 1998: *From Metaphysics to Ethics*. Oxford: Oxford University Press.

Kim, J. 1984: "Concepts of Supervenience". Reprinted in Kim 1993. All page references to this text.

Kim, J. 1987: "'Strong' and 'Global' Supervenience Revisited". Reprinted in Kim 1993. All page references to this text.

Kim, J. 1990: "Supervenience as a Philosophical Concept". Reprinted in Kim 1993. All page references to this text.

Kim, J. 1993: *Supervenience and Mind*. Cambridge: Cambridge University Press.

Kim, J. 1996: *Philosophy of Mind*. Oxford: Westview Press.

Kim, J. 2000: *Mind in a Physical World*. The MIT Press: Cambridge Massachusetts.

Klagge, J. 1984: "An Alleged Difficulty Concerning Moral Properties". *Mind* Vol. XCIII.

Klagge, J. 1988: "Supervenience: Ontological and Ascriptive". *Australasian Journal of Philosophy*, Vol 66.

Kripke, S. 1980: *Naming and Necessity*. Cambridge, MA: Harvard University Press.

- Lewis, D. 1986: *On the Plurality of Worlds*. Oxford: Basil Blackwell.
- Lovibond, S. 1983: *Realism and Imagination in Ethics*. Oxford: Blackwell.
- Mackie, J. 1977: *Ethics: Inventing Right and Wrong*. New York: Penguin.
- McDowell, J. 1979: "Virtue and Reason". *The Monist* 63, 331-43.
- McDowell, J. 1985: "Values and Secondary Qualities". Honderich, T. (ed.), *Morality and Objectivity: Essays in Honour of J. L. Mackie*. London: Routledge & Kegan Paul.
- McDowell, J. 1995: "Two Sorts of Naturalism". Hursthouse, Lawrence and Quinn (eds.), *Virtues and Reasons*. Oxford: Oxford University Press.
- McFettridge, I. G. 1990: *Logical Necessity and Other Essays*. Haldane, J. & Scruton, R. (eds.). *The Aristotelian Society Series*, Vol. 11. London: The Aristotelian Society.
- McLaughlin, B. 1995: "Varieties of Supervenience". Savellos & Yalcin (eds.), *Supervenience: New Essays*. Cambridge: Cambridge University Press.
- McNaughton, D. 1988: *Moral Vision*. Oxford: Blackwell.
- Miller, A. 1998: *Philosophy of Language*. London: UCL Press.
- Miller, A. 2003: *An Introduction to Contemporary Metaethics*. Polity Press.
- Moore, G. E. [1903] 1993: *Principia Ethica*, rev. edn. Cambridge: Cambridge University Press.
- Moore, G. E. 1942: "A Reply to My Critics". Schilpp, P.A. (ed.), *The Philosophy of G. E. Moore*. Illinois: Chicago and Evanston.

Moser, P. & Yandell, D. 2000: "Farewell to Philosophical Naturalism". Craig, W. & Moreland, J. (eds.), *Naturalism*. London: Routledge Press.

Papineau, D. 1993: *Philosophical Naturalism*. Oxford: Blackwell Press.

Pigden, C. 1993: "Naturalism". Singer, P. (ed.), *A Companion to Ethics*. Oxford: Blackwell Press.

Platts, M. 1979: *The Ways of Meaning*. London: Routledge & Kegan Paul.

Priest, G. 2001: *An Introduction to Non-Classical Logic*. Cambridge: Cambridge University Press.

Prior, A. N. 1960: "The Runabout Inference-Ticket". *Analysis* 21, 38-9.

Putnam, H. 1967: "The Mental Life of Some Machines". Castaneda, H. (ed.), *Intentionality, Minds and Perception*. Detroit: Wayne State University Press. Reprinted in Putnam 1975b. All page references to this text.

Putnam, H. 1975a: "The Meaning of 'Meaning'". Gundersen, K. (ed.), *Language, Mind and Knowledge*. Minnesota Studies in The Philosophy of Science VII. Minnesota: Minnesota University Press. Reprinted in Putnam 1975b. All page references to this text.

Putnam, H. 1975b: *Mind, Language, and Reality*. Cambridge: Cambridge University Press.

Rachels, J. 2000: "Naturalism". LaFolette, H. (ed.), *The Blackwell Guide to Ethical Theory*. Oxford: Blackwell Press.

Railton, P. 1986: "Moral Realism". *Philosophical Review* 95, 163-207.

Railton, P. 1989: "Naturalism and Prescriptivity". *Social Philosophy and Policy* 7, 151-74.

Railton, P. 1993a. "What the Non-Cognitivist Helps Us To See the Naturalist Must Help Us To Explain". Haldane, J. & Wright (eds.) *Reality, Representation, and Projection*. Oxford: Oxford University Press, 279-300.

Railton, P. 1993b: "Reply to David Wiggins". Haldane, J. & Wright, C. (eds.), *Reality, Representation, and Projection*. Oxford: Oxford University Press, 315-28.

Railton, P. 1995: "Made In the Shade: Moral Compatibilism and the Aims of Moral theory". Couture, J & Nielsen, K. (eds.), *On the Relevance of Metaethics: New Essays on Metaethics*. Canadian Journal of Philosophy. Calgary: University of Calgary Press.

Sayre-McCord, G. 1986: "The Many Moral Realisms". *Southern Journal of Philosophy* 24, suppl., 1-22.

Schueler, G. F. 1988: "Modus Ponens and Moral Realism". *Ethics* 98, 501-517.

Schwartz, S. P. (ed.) 1977: *Naming, Necessity and Natural Kinds*. Ithaca: Cornell University Press.

Singer, P. (ed.) 1991: *A Companion to Ethics*. Oxford: Blackwell.

Smith, M. 1994: *The Moral Problem*. Oxford: Blackwell.

Smith, M. 2000: "Moral Realism". LaFolette, H. (ed.), *The Blackwell Guide to Ethical Theory*. Oxford: Blackwell Press.

Skorupski, J. 1999: "Irrealist Cognitivism". *Ratio*, Vol. XII.

Sturgeon, N. 1988: "Moral Explanations". Sayre-McCord, G. (ed.), *Essays on Moral Realism*. Ithaca, NY: Cornell University Press, 229-55.

Timmons, M. 1999: *Morality without Foundations*. Oxford: Oxford University Press.

Unwin, N. 1999: "Quasi-realism, Negation and the Frege-Geach Point". *Philosophical Quarterly*, vol. 49.

Van Roojen, M. 1996: "Expressivism and Irrationality". *Philosophical Review* 105, 311-35.

Wagner, S. & Warner, R. (eds.) 1993: *Naturalism*. Indiana: University of Notre Dame Press.

Wright, C. 1985: "Review of Blackburn 1984". *Mind* 94, 310-19.

Wright, C. 1988: "Realism, Antirealism, Irrealism, Quasi-realism". *Midwest Studies in Philosophy* 12, 25-49.

Wright, C. 1996: "Truth in Ethics". Hooker, B. (ed.), *Truth in Ethics*. Oxford: Blackwell, 1-18.

Zangwill, N. 1992: "Moral Modus Ponens". *Ratio* 5, 177-93.