



National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file *Votre référence*

Our file *Notre référence*

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

A NEW THEORY OF NONCOGNITIVISM

by

Wayne Fenske

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
August, 1995

© Copyright by Wayne Fenske, 1995.



National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file *Voire référence*

Our file *Notre référence*

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-612-08761-1

Canada

Name Wayne Fenske

Dissertation Abstracts International is arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation. Enter the corresponding four digit code in the spaces provided.

Philosophy

0422

U·M·I

SUBJECT TERM

SUBJECT CODE

Subject Categories

THE HUMANITIES AND SOCIAL SCIENCES

COMMUNICATIONS AND THE ARTS

Architecture 0729
 Art History 0377
 Cinema 0900
 Dance 0378
 Fine Arts 0257
 Information Science 0723
 Journalism 0391
 Library Science 0399
 Mass Communications 0708
 Music 0413
 Speech Communication 0459
 Theater 0465

EDUCATION

General 0515
 Administration 0514
 Adult and Continuing 0516
 Agricultural 0517
 Art 0273
 Bilingual and Multicultural 0282
 Business 0688
 Community College 0275
 Curriculum and Instruction 0727
 Early Childhood 0518
 Elementary 0524
 Finance 0277
 Guidance and Counseling 0519
 Health 0680
 Higher 0745
 History of 0520
 Home Economics 0278
 Industrial 0521
 Language and Literature 0279
 Mathematics 0280
 Music 0522
 Philosophy of 0998
 Physical 0523

Psychology 0525
 Reading 0535
 Religious 0527
 Sciences 0714
 Secondary 0533
 Social Sciences 0534
 Sociology of 0340
 Special 0529
 Teacher Training 0530
 Technology 0710
 Tests and Measurements 0288
 Vocational 0747

LANGUAGE, LITERATURE AND LINGUISTICS

Language 0679
 General 0289
 Ancient 0290
 Linguistics 0291
 Modern 0401
 Literature 0294
 General 0295
 Classical 0297
 Comparative 0298
 Medieval 0316
 Modern 0591
 African 0305
 American 0352
 Asian 0355
 Canadian (English) 0593
 Canadian (French) 0311
 English 0312
 Germanic 0315
 Latin American 0313
 Middle Eastern 0314
 Romance 0578
 Slavic and East European

PHILOSOPHY, RELIGION AND THEOLOGY

Philosophy 0422
 Religion 0318
 General 0321
 Biblical Studies 0319
 Clergy 0320
 History of 0322
 Philosophy of 0469
 Theology

SOCIAL SCIENCES

American Studies 0323
 Anthropology 0324
 Archaeology 0326
 Cultural 0327
 Physical 0310
 Business Administration 0272
 General 0770
 Accounting 0454
 Banking 0338
 Management 0385
 Marketing 0501
 Canadian Studies 0503
 Economics 0505
 General 0508
 Agricultural 0509
 Commerce Business 0510
 Finance 0511
 History 0358
 Labor 0366
 Theory 0351
 Folklore 0578
 Geography
 Gerontology
 History
 General

Ancient 0579
 Medieval 0581
 Modern 0582
 Black 0328
 African 0331
 Asia, Australia and Oceania 0332
 Canadian 0334
 European 0335
 Latin American 0336
 Middle Eastern 0333
 United States 0585
 History of Science 0398
 Law
 Political Science 0615
 General
 International Law and Relations 0616
 Public Administration 0617
 Recreation 0814
 Social Work 0452
 Sociology 0626
 General
 Criminology and Penology 0627
 Demography 0938
 Ethnic and Racial Studies 0631
 Individual and Family Studies 0628
 Industrial and Labor Relations 0629
 Public and Social Welfare 0630
 Social Structure and Development 0700
 Theory and Methods 0344
 Transportation 0709
 Urban and Regional Planning 0999
 Women's Studies 0453

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES

Agriculture 0473
 General 0285
 Agronomy
 Animal Culture and Nutrition 0475
 Animal Pathology 0476
 Food Science and Technology 0359
 Forestry and Wildlife 0478
 Plant Culture 0479
 Plant Pathology 0480
 Plant Physiology 0917
 Range Management 0777
 Wood Technology 0746
 Biology 0306
 General 0287
 Anatomy 0308
 Biostatistics 0309
 Botany 0379
 Cell 0329
 Ecology 0353
 Entomology 0369
 Genetics 0793
 Limnology 0410
 Microbiology 0307
 Molecular 0317
 Neuroscience 0416
 Oceanography 0433
 Physiology 0821
 Radiation 0778
 Veterinary Science 0472
 Zoology
 Biophysics 0786
 General 0760
 Medical
 Earth Sciences 0425
 Biogeochemistry 0996
 Geochemistry

Geodesy 0370
 Geology 0372
 Geophysics 0373
 Hydrology 0388
 Mineralogy 0411
 Paleobotany 0345
 Paleoecology 0426
 Paleontology 0418
 Paleozoology 0985
 Palynology 0427
 Physical Geography 0368
 Physical Oceanography 0415

HEALTH AND ENVIRONMENTAL SCIENCES

Environmental Sciences 0768
 Health Sciences 0566
 General 0300
 Audiology 0992
 Chemotherapy 0567
 Dentistry 0350
 Education 0769
 Hospital Management 0758
 Human Development 0982
 Immunology 0564
 Medicine and Surgery 0347
 Mental Health 0569
 Nursing 0570
 Nutrition 0380
 Obstetrics and Gynecology 0354
 Occupational Health and Therapy 0381
 Ophthalmology 0571
 Pathology 0419
 Pharmacology 0572
 Pharmacy 0382
 Physical Therapy 0573
 Public Health 0574
 Radiology 0575
 Recreation

Speech Pathology 0460
 Toxicology 0383
 Home Economics 0386

PHYSICAL SCIENCES

Pure Sciences
 Chemistry 0485
 General 0749
 Agricultural 0486
 Analytical 0487
 Biochemistry 0488
 Inorganic 0738
 Nuclear 0490
 Organic 0491
 Pharmaceutical 0494
 Physical 0495
 Polymer 0754
 Radiation 0405
 Mathematics 0605
 Physics 0986
 General
 Acoustics
 Astronomy and Astrophysics 0606
 Atmospheric Science 0608
 Atomic 0748
 Electronics and Electricity 0607
 Elementary Particles and High Energy 0798
 Fluid and Plasma 0759
 Molecular 0609
 Nuclear 0610
 Optics 0752
 Radiation 0753
 Solid State 0611
 Statistics 0463
 Applied Sciences
 Applied Mechanics 0346
 Computer Science 0984

Engineering 0537
 General 0538
 Aerospace 0539
 Agricultural 0540
 Automotive 0541
 Biomedical 0542
 Chemical 0543
 Civil 0544
 Electronics and Electrical 0348
 Heat and Thermodynamics 0545
 Hydraulic 0546
 Industrial 0547
 Marine 0794
 Materials Science 0548
 Mechanical 0743
 Metallurgy 0551
 Mining 0552
 Nuclear 0549
 Packaging 0765
 Petroleum 0554
 Sanitary and Municipal System Science 0790
 Geotechnology 0428
 Operations Research 0796
 Plastics Technology 0795
 Textile Technology 0994

PSYCHOLOGY

General 0621
 Behavioral 0384
 Clinical 0622
 Developmental 0620
 Experimental 0623
 Industrial 0624
 Personality 0625
 Physiological 0989
 Psychobiology 0349
 Psychometrics 0632
 Social 0451



Table Of Contents

Abstract.....	p.vi
---------------	------

Chapter One: Moral Twin-Earth

1.1: Introduction.....	p.1
1.2: Moore's Argument Against Moral Naturalism.....	p.3
1.3: Frankena's Criticism Of Moore.....	p.11
1.4: Moore's Theoretical Background.....	p.12
1.5: Synonymy.....	p.17
1.6: Can Moore's Argument Be Salvaged?.....	p.19
1.7: Kripke's Theory Of Reference.....	p.21
1.8: Putnam's Twin Worlds.....	p.28
1.9: Moore's Open Question Argument Rendered Obsolete.....	p.31
1.10: An Updated Semantics For Moral Terminology.....	p.34
1.11: Supervenience Accounts Of Moral Properties.....	p.38
1.12: Dispositional Accounts Of Moral Properties.....	p.43
1.13: An Initial Problem For Updated Moral Semantics.....	p.46
1.14: Moral Twin-Earth.....	p.51
1.15: Disagreements Over The Nature Of Justice.....	p.53
1.16: Potential Problems With Moral Twin-Earth.....	p.62

Chapter Two: The Argument From Internalism

2.1: Introduction.....	p.73
2.2: A Three-Level Version Of Noncognitivism.....	p.74
2.3: Two Distinctions Pertaining To Internalism.....	p.79
2.4: Brink's Three-Step Characterization Of Internalism...p.85	
2.5: Internalism, Noncognitivism, And The Amoralist.....	p.86
2.6: The 'Oh Yeah?' Debate.....	p.89
2.7: A Crude View Of Conceptual Truth.....	p.95
2.8: Conceptual Truth And Moore's Original OQA.....	p.102
2.9: Two Kinds Of Necessary Connection.....	p.104
2.10: Moral Twin-Earth Reconsidered.....	p.107
2.11: Moral Language Use As Essentially Group-Oriented...p.113	
2.12: The Deviant Chess Player.....	p.118
2.13: The Amoralist As Deviant Moral Language User.....	p.123
2.14: Williams' Integrity Objection.....	p.131
2.15: A Problem For The Externalist.....	p.138
2.16: A Second Problem For The Externalist.....	p.143

Chapter Three: The Embedding Problem

3.1: Introduction.....	p.148
3.2: Geach And 'The Frege Point'.....	p.150
3.3: Searle's Parallel Criticism.....	p.153
3.4: The Target: R.M. Hare's Noncognitivism.....	p.154
3.5: Hare's Dilemma.....	p.157
3.6: Noncognitivism As An A Posteriori Doctrine.....	p.159
3.7: Gibbard's Solution To The Embedding Problem.....	p.161
3.8: Elaborating On Some Of Gibbard's Insights.....	p.167
3.9: Validity And Normativity: Two Separate Issues.....	p.175

Chapter Four: Moral Explanations

4.1: Introduction.....	p.179
4.2: Epistemology Naturalized In Morality And Science....	p.183
4.3: Moral Explanations.....	p.187
4.4: Explanatory Relevance And Explanatory Potence.....	p.190
4.5: Supervenience And Necessity.....	p.192
4.6: Six Aspects Of Comparison.....	p.194
4.7: True Moral Theories.....	p.202
4.8: The Second And Fifth Explanans Reconsidered.....	p.210
4.9: Strong Moral Properties.....	p.219
4.10: Justified Standards And Ideal Conditions.....	p.224
4.11: Tar-Maintaining.....	p.226
4.12: Normative Presuppositions And Ideal Situations....	p.229
4.13: Tar-Maintaining Reconsidered.....	p.234
4.14: Irreconcilable Differences.....	p.238
4.15: Morally Relevant Differences.....	p.243
4.16: Concluding Remarks.....	p.248

Bibliography.....	p.251
-------------------	-------

ABSTRACT

In the dissertation, I attempt to defend a new theory of noncognitivism. I do this by discussing four issues, three of which pertain to the views of a group of scholars who have come to be known as 'The Cornell Moral Realists'. One of the most impressive aspects of the Cornell metaethics is its theoretical expansiveness. It encompasses views in philosophy of language, philosophy of science, philosophical psychology, as well as normative ethics. As such, it serves as a useful foil against which to contrast my own metaethical position. In the first chapter, I begin by discussing the viability of G.E. Moore's Open Question Argument. I contend that an updated version of Moore's OQA allows me to establish a preliminary argument in favor of noncognitivism. I begin the second chapter by explicating more fully the nature of the particular version of noncognitivism I espouse. From there, I go on to discuss the question of whether or not the amoralist is conceivable. Rather than arguing that the amoralist is not conceivable, I explain how the conceivability of the amoralist does not present a problem for the version of noncognitivism I am espousing. In the third chapter, I explain how to construct a noncognitivist semantics that can be used to undergird valid patterns of inference which contain noncognitive constituents. The third chapter is unique in its departure from direct discussion of The Cornell Moral Realists. In the final chapter, the focus of my campaign is Nicholas Sturgeon's influential discussion of the possibility of moral explanations. I contend that the existence of moral properties cannot be vindicated by attempting to parallel moral inquiry with scientific inquiry. The discussion of these four issues provides a synoptic defense of a new theory of noncognitivism.

CHAPTER ONE: MORAL TWIN-EARTH

1.1: Introduction

Consider the following question: What are you doing when you make a moral utterance like 'Stealing is wrong'? The question itself appears to be simple enough. The answer will not be, however. In fact, I will spend this entire dissertation attempting to defend a particular kind of answer to this question. In doing so, I will set out what I take to be the truth but not the whole truth on this issue. This is a matter of practical necessity. The literature in this area is vast. In a recent review article, Darwell, Gibbard, and Railton spend 75 pages simply providing cursory sketches of some of the different metaethical stances that have been espoused by various authors.¹ We will have to start somewhere. As Margaret Little points out in her marvellous pair of review articles on contemporary literature in metaethics:

There is a deep and fascinating divide between the naturalist moral realism popular in American philosophical circles and the non-naturalist theory prevailing in the United Kingdom.²

She describes the crucial difference between the approaches adopted by the two camps in the following manner.

While American moral realists have, by and large, depicted moral theory as continuous with scientific theory, recent English moral realists have insisted on certain fundamental ways in which morality differs from

¹Darwell, Gibbard, and Railton, 1992, pp.115-189.

²Little, 1994a, p.145.

science, rejecting the idea that such differences constitute deficiencies.³

This divide provides me with the starting point that will allow me to focus my answer to this question.

In what follows, I will confine myself to pursuing a set of issues on the American side of this debate. In particular, I will be principally concerned with discussing the views of a group of scholars who have come to be known as 'The Cornell Moral Realists'. One of the most impressive aspects of the Cornell metaethics is its theoretical expansiveness. It encompasses views in philosophy of language, philosophy of science, philosophical psychology, as well as normative ethics. As such, it will serve as a useful foil against which to contrast my own metaethical position. In the first chapter, I will begin by discussing a metaethical question which pertains to issues in the philosophy of language. The question concerns the viability of G.E. Moore's Open Question Argument. I contend that an updated version of Moore's OQA allows me to establish a preliminary argument in favor of noncognitivism. I will begin the second chapter by explicating the nature of the particular version of noncognitivism I espouse more fully. From there, I will go on to discuss a metaethical question which pertains to an issue in philosophical psychology. This is the question of whether or not the amoralist is conceivable. I contend that the amoralist does not present a

³Little, 1994b, p.225.

problem for the version of noncognitivism I am espousing. In the third chapter, I will discuss a metaethical question which pertains to issues in logic and proof theory. I will explain how to construct a noncognitivist semantics that will undergird valid patterns of inference which contain noncognitive constituents. The third chapter will be unique in its departure from direct discussion of The Cornell Moral Realists. In the final chapter, I will discuss a metaethical question which pertains to issues in the philosophy of science. The focus of my campaign will be Nicholas Sturgeon's influential discussion of the possibility of moral explanations. I contend that the existence of moral properties cannot be vindicated by attempting to parallel moral inquiry with scientific inquiry. Together, the discussion of these four issues will provide a synoptic defense of a new theory of noncognitivism.

1.2: Moore's Argument Against Moral Naturalism

I can begin my case for a new theory of noncognitivism by considering the historical influence of G.E. Moore. Since the time of G.E. Moore, the existence of objective moral facts has largely been rejected by the analytic tradition. This rejection is based on the belief that it is not possible to accommodate moral facts within a naturalistic framework. Moore rejected the possibility of ethical naturalism in favor of a non-naturalist ethical realism where moral properties and facts are part of a sui generis non-natural realm. He was also

willing to claim that human beings are equipped with special intuitive abilities which enable us to detect these sui generis moral phenomena. According to Moore's model, scientific enquiry and moral enquiry are entirely distinct ways of obtaining knowledge from entirely distinct realms. Regardless of the flaws with Moore's position, which were known relatively early, **Principia Ethica** was still influential in establishing the predominance of the view that there is a sharp divide between the discourse of science and the discourse of morality. The fact-value distinction was regarded as a truism by virtually everyone. Not surprisingly, people soon began to reject Moore's intuitionistic realism as an adequate characterization of morality. They did not reject it in favor of a naturalistic moral realism, however. Many turned instead to the view that moral language is principally used to express attitudes rather than to describe states of affairs. Thus, noncognitivism was the principal historical beneficiary of Moore's arguments.

In order to assess the relative merits of Moore's position, it will be useful to consider his claims about the nature of moral goodness in detail. Moore begins **Principia Ethica** with an attempt to provide an adequate characterization of the subject matter of his book. He considers the possibility that Ethics is that area of study which is concerned with the proper regulation of human conduct, but finds this characterization too narrow since it seems obvious

that Ethics is concerned with other matters besides the regulation of conduct. For Moore, Ethics is a discursive activity which is specifically concerned with the application of the predicate 'good' (and its converse 'bad'). This predicate can be applied to the entire range of moral assessment which includes persons, actions, and states of affairs. In order to understand the nature of Ethics more clearly, Moore wants to focus on the task of determining what this central predicate **means**. This will enable us to ascertain exactly what we are doing when we make Ethical utterances. He is engaged in an analysis of an important aspect of our linguistic practice here. Moore felt that his predecessors in moral philosophy failed to understand the nature of Ethics because they did not understand the meaning of its central predicate 'good'. Thus, it was of utmost importance to provide an adequate analysis of this crucial Ethical concept. According to Moore:

this question, how 'good' is to be defined, is the most fundamental question in all Ethics. That which is meant by 'good' is, in fact, except for its converse 'bad' the **only** simple object of thought which is peculiar to Ethics. Its definition is, therefore, the most essential point in the definition of Ethics;⁴

Having argued that ascertaining the meaning of the word 'good' is the essential first step in arriving at an adequate understanding of the subject of Ethics, Moore goes on to claim that this all-important word is, in fact, indefinable. The

⁴Moore, 1903, p.5.

word 'good' denotes a simple, unanalyzable concept. As such, we cannot explain the nature of goodness to anyone who does not already understand what it is. In his words:

My point is that 'good' is a simple notion, just as 'yellow' is a simple notion; that, just as you cannot, by any manner of means, explain to anyone who does not already know it, what yellow is, so you cannot explain what good is.⁵

In his mind, a major problem with moral philosophy is that people have been unwilling to take goodness for the simple, unanalyzable concept that it is. Instead, they have treated it as if it were a complex notion by attempting to provide an analysis of it. To engage in such an enterprise is to misunderstand the nature of goodness. It is to commit the naturalistic fallacy. He felt that a number of thinkers were doing precisely this when they attempted to describe goodness in terms of collections of natural properties. He claimed that:

far too many philosophers have thought that when they named those other properties they were actually defining good; that these properties, in fact, were simply not 'other', but absolutely and entirely the same with goodness. This view I propose to call the 'naturalistic fallacy'...⁶

Moore goes on to attempt to demonstrate why this approach to formulating a definition of the notion of goodness should be considered to be fallacious. The general type of fallacy of which the naturalistic fallacy is a specific variety occurs

⁵Moore, 1903, p.7.

⁶Moore, 1903, p.10.

whenever anyone unwittingly conflates the 'is' of predication with the 'is' of identity. In order to illustrate the generic fallacy involved here, Moore offers the following example:

If I were to imagine that when I said 'I am pleased', I meant that I was exactly the same thing as 'pleased', I should not indeed call that a naturalistic fallacy, although it would be the same fallacy as I have called naturalistic with reference to Ethics.⁷

The reason it would not be appropriate to call this particular fallacy **naturalistic** is that it involves the mistaken conflation of two objects (i.e., the person in question and pleasure) which are both part of the natural world. When good, however, is mistakenly taken to be identical to some class of natural states of affairs, there is reason to refer to this as the **naturalistic** fallacy. To do this is to conflate an aspect of an entirely distinct sui generis realm with some aspect of the natural world. It is to mistakenly think that something which is not natural is, in fact, natural. It is to disregard Bishop Butler's motto "Everything is what it is, and not another thing" - the slogan Moore placed on the title page of **Principia Ethica**. Speaking of someone who commits this fallacy, Moore says:

if he confuses 'good', which is not in the same sense a natural object, with any natural object whatever, then there is a reason for calling that a naturalistic fallacy; its being made with regard to 'good' marks it as something quite specific, and this specific mistake deserves a name because it is so common.⁸

⁷Moore, 1903, p.13.

⁸Moore, 1903, p.13.

The problem with many moral theorists, according to Moore, is that in saying things about that which is good, they mistakenly thought that they were providing an analysis of the meaning of 'good'. In doing so, they conflated the 'is' of predication with the 'is' of identity and committed the naturalistic fallacy.

Moore now moves on to present his positive argument for the claim that 'good' is indefinable. He attempts to convince the reader by using a trilemma strategy. The alternatives are that 'good' is either indefinable, complex, or has no meaning at all. By showing that the last two alternatives are implausible, he will thus have shown that 'good' is indefinable. In order to accomplish this task, he adduces a single argument which is able to dispose of both of them. This has come to be known as 'The Open Question Argument'. He eliminates the possibility that 'good' might be a complex notion by considering what a disagreement about the meaning of 'good' might amount to. He says:

The hypothesis that disagreement about the meaning of good is disagreement with regard to a correct analysis of a given whole, may be most plainly seen to be incorrect by consideration of the fact that, whatever definition be offered, it may always be asked, with significance, of the complex so defined, whether it is itself good.⁹

He eliminates the possibility that 'good' might have no meaning by considering what would have to be the case for us

⁹Moore, 1903, p.15.

to be able to doubt whether any particular natural state of affairs is actually good. In his words:

whoever will attentively consider with himself what is actually before his mind when he asks the question 'Is pleasure (or whatever it may be) after all good?' can easily satisfy himself that he is not merely wondering whether pleasure is pleasant. And if he will try this experiment with each suggested definition in succession, he may become expert enough to recognize that in every case he has before his mind a unique object, with regard to the connection of which with any other object, a distinct question may be asked. Everyone does in fact understand the question 'Is this good?' When he thinks of it, his state of mind is different from what it would be, were he asked 'Is this pleasant, or desired, or approved?' It has a distinct meaning for him, even though he may not recognize in what respect it is distinct.¹⁰

That is to say, the person must already have a distinct concept of 'good' in her mind in order to be able to wonder whether or not it is correct to predicate that concept of pleasure or any other quality or state of affairs which is purported to be good. Since the Open Question Argument shows that 'good' cannot be complex or meaningless, we are forced to conclude that it is simple and indefinable.

Putnam provides an alternative rendering of the essence of the Open Question Argument by considering the status of a statement which claims that some proffered analysis of goodness **lacks** rather than possesses the feature of goodness. If the analysis of 'good' contained in this statement is accepted as correct, then the statement will be false by definition rather than true by definition as in Moore's original formulation. As such, it will not be an open question

¹⁰Moore, 1903, pp.16,17.

whether or not the statement could possibly be true. Yet, even the person who accepted the analysis of 'good' contained in that statement would not take the statement to be false by definition. Thus, no proffered analysis of the nature of goodness could serve as an explication of the meaning of 'good'. In Putnam's words:

Moore's argument that Good cannot be a physicalistic property (a 'natural' property) was that if 'Good' is the same property as 'conducive to maximizing total utility' (or whatever natural property, physical or functional, you care to substitute), then

(1) 'this action is not good even though it is conducive to maximizing total utility' is a **self-contradictory** statement (not just a **false** one).

But even a Utilitarian would not claim (1) is **self-contradictory**. And this shows, Moore claims, that although being Good and being conducive to maximizing total utility might be **correlated** properties, they could not be the same **identical** property.¹¹

A crucial passage in *Principia Ethica* clearly illustrates that one of the motivations behind Moore's formulation of the Open Question Argument is that he was trying to preserve the intelligibility of ethical disagreement. For him, the problem with attempting to formulate a definition of 'good' in terms of natural qualities is that this approach cuts off the possibility of legitimate ethical debate in advance. This is the reason we must reject all attempts to provide naturalistic definitions of goodness. As he says:

If we start with the conviction that a definition of good can be found, we start with the conviction that good **can mean** nothing else than some one property of things; and our only business will then be to discover what that property is. ... For we shall start with the conviction

¹¹Putnam, 1981, p.206.

that good must mean so and so, and shall therefore be inclined either to misunderstand our opponent's arguments or to cut them short with the reply, 'This is not an open question: the very meaning of the word decides it; no one can think otherwise except through confusion.'¹²

1.3: Frankena's Criticism Of Moore

In his influential article, "The Naturalistic Fallacy.", Frankena claims that the particular mistake Moore is attributing to these moral theorists has nothing to do with the putative fact that they are trying to interdefine entities from distinct ontological realms. This observation is often overlooked by those whose interest is specifically focused on the issue of whether or not Moore's arguments provide any buttressing for the claim that it is not possible to validly derive an 'ought' statement from an 'is' statement. Frankena proposes that we concentrate on discerning the nature of the generic fallacy of which the naturalistic fallacy is but an instance in order to home in on the root of Moore's criticism. He calls this generic fallacy 'The Definist Fallacy' which he describes as follows:

the definist fallacy is the process of confusing or identifying two properties, of defining one property by another, or of substituting one property for another. Furthermore, the fallacy is always simply that two properties are being treated as one, and it is irrelevant, if it be the case, that one of them is natural or non-ethical and the other non-natural or ethical.¹³

¹²Moore, 1903, pp.20,21.

¹³Frankena, 1939, p.471.

This is what ethical theorists are doing when they attempt to formulate naturalistic definitions of 'good'. The problem is not that they have selected unsuitable types of properties as the material from which they attempt to fashion the essential nature of goodness. The problem is just that they fail to see that any property or collection of properties they choose will still be distinct from the property which constitutes goodness. To treat these two distinct properties as if they were one and the same property is to commit the definist fallacy. Once this is understood, however, it quickly becomes obvious that Moore's argument as it stands shows nothing peculiar about the nature of goodness. If the definist fallacy constitutes a legitimate criticism of all attempts to provide a suitable analysis of the meaning of 'good', then it can also be used to criticize all attempts to provide suitable analyses of all kinds of other concepts as well. These will also turn out to be mistaken conflations of two distinct properties or sets of properties. As Frankena claims:

if Mr. Moore's motto (or the definist fallacy) rules out any definitions, for example of 'good', then it rules out all definitions of any term whatever.¹⁴

1.4: Moore's Theoretical Background

It seems to me that Frankena's discussion does not do justice to the point that Moore is attempting to make with the Open Question Argument. It is significant to notice that Moore did not accept the Open Question Argument as a reliable

¹⁴Frankena, 1939, p.472.

generic strategy that could be used to illustrate instances of the 'definist fallacy'. He was ready to admit that allowing the Open Question Argument to be used as a standard for testing the validity of definitions in general would lead to the anti-intuitive consequences he called 'the paradox of analysis'.¹⁵ If all legitimate analyses had to conform to the dictates of the Open Question Argument, it would be impossible to formulate an acceptable definition of any term. It is important to keep in mind that what Moore was attempting to do with the Open Question Argument was to convince the reader that there is something special about the nature of goodness in particular. In presenting his formulation of it, he assumed that his views on semantics and epistemology were essentially correct. It is significant to notice that his views on ethics informed his views on semantics and epistemology and not vice versa. Moore's earliest extended philosophical studies were two dissertations written in 1897 and 1898 respectively. Each of these were entitled **The Metaphysical Basis Of Ethics**. He won a prize fellowship for the second dissertation which allowed him to stay on at Cambridge for an additional six years. The second dissertation provides the first detailed description of the semantic and epistemological theory which underlies the reasoning contained in the Open Question Argument.

Alan R. White referred to Moore's semantics as:

¹⁵Darwell, Gibbard, Railton, 1992, p.115.

the **concept theory** of meaning, for it is the view that the meaning of an expression is the concept - or the proposition, as it is called when the expression is a complete sentence - which the expression stands for, signifies, names or expresses.¹⁶

According to this characterization of semantics, the meaning of an utterance is determined by ascertaining the intentions of the utterer. The meaning of a term or expression is the particular concept that speakers intend to express by using that term or expression. The meaning of a sentence is the particular proposition speakers intend to express by using that sentence. Speaking of propositions in particular, Moore claimed that:

a proposition, is **not** a name for any mere form of words. It is a name for what is **expressed** by certain forms of words - those, namely, which, in grammar, are called 'sentences'. It is a name for what is before your mind, when you not only hear or read but **understand** a sentence. It is, in short, the **meaning** of a sentence - what is expressed or conveyed by a sentence: and is, therefore, utterly different from the sentence itself - from the mere words.¹⁷

It is the object of belief that the mind is focused upon when it apprehends the meaning of a sentence. Moore held a parallel view of the meanings of parts of sentences. Their meanings are constituted by the concepts which they express. Tom Regan, citing from an unpublished manuscript of Moore's, tells us that:

¹⁶White, 1958, p.40.

¹⁷Moore, 1910, p.259.

"A concept", Moore asserts, "is not a mental fact nor any part of a mental fact"(p.179). Rather, it is **what** we mean when, for example, we use the words 'red' and 'rose'.¹⁸

It is important to notice that Moore does not see the concepts which are expressed by language as mere mental entities. They are universal entities which exist independently of human minds, much like Platonic Forms. They play an important role in human thought processes.

in order to make judgments, Moore maintains that we must grasp these concepts themselves. We must have a **direct** grasp of the concepts Red and Rose, for example, to make the judgment "This rose is red." When we think this thought, or make this judgment, it is not **our ideas** of red and rose that we are thinking of; it is these concepts **themselves**.¹⁹

Moore's understanding of synonymy is closely tied to his concept theory of meaning. White describes the connection between them as follows:

to say that two expressions are synonymous is to say either that the concept or proposition expressed by one is the same as the concept or proposition expressed by the other or that both expressions express one and the same concept or proposition.²⁰

Thus, two expressions mean the same thing only if they bring the same concept to mind. As Regan notes, Moore applies the concept theory of meaning to his discussion of moral vocabulary as well.

The meaning of 'red' is the concept Red; of 'rose' the concept Rose; and so on. There is in every case only one universal meaning though there may be many instances -

¹⁸Regan, 1986, p.104.

¹⁹Regan, 1986, p.107.

²⁰White, 1958, p.98.

many red things, but only one concept, Red; many roses, but only one concept Rose; and so on. Moore in **Principia** continues to operate with this ontology of meanings in the background, so that when he insists that the **object** of Ethics is Good, he means that Ethics must start with an examination of this particular concept, object, notion, or, in other words, this particular universal meaning. Accordingly, by claiming that Good is indefinable Moore does not mean that the English word 'good' lacks a dictionary definition. He means that the universal meaning of the word 'good' - the concept, notion, object, or idea that this word, or its equivalent in any natural language, names, denotes, or signifies - cannot be defined. The **notion** Good is not identical with any notion other than itself.²¹

If the word 'good' were synonymous with any expression which describes some natural state of affairs, then both expressions would designate one and the same concept in the mind of anyone who was considering the two expressions. But if this were so, it would not be possible for her to consider whether or not the concept corresponding to the description of the natural state of affairs was the same as the concept corresponding to the word 'good' since both expressions would evoke the same concept in the mind. Implicit in this argument is the presumption that concepts are of such a transparent nature that it is not possible for a cognizer to fail to recognize those cases where two expressions turn out to refer to the same concept. Concepts, rather than words, are the medium of thought and it will always be obvious to the cognizer when she is using two expressions to bring one and the same concept before her mind. In such cases, she could hardly miss the fact that she is thinking the same thing in both instances. Since

²¹Regan, 1986, pp.193,194.

she couldn't miss this, it obviously couldn't be an open question for her whether or not the two expressions refer to the same concept and, thus, have the same meaning. This is the key semantic presupposition underlying Moore's Open Question Argument.

1.5: Synonymy

The major difficulty that the early critics found with Moore's argument against ethical naturalism is that it presupposes a synonymy criterion of property identity. That is, it presupposes that if property x is identical to property y , then the criteria one requires for the presence of x will be the same as the criteria one requires for the presence of y . So, if x is identical to y , and one concedes that the requisite criteria for the presence of x have been satisfied, one is conceptually committed to conceding that the requisite criteria for the presence of y have also been satisfied. Thus, once one has become epistemically committed to the presence of x , whether or not y is also present can no longer be an open question for her. Y has to be present because the requisite criteria for its presence have admittedly already been satisfied.

The crucial problem with Moore's synonymy criterion of property identity is that it presupposes that all of the concepts we use are referentially transparent. When a concept is referentially transparent, the user of that concept is fully aware of its exact extension. In fact, however, we often

use concepts which are referentially opaque. When a concept is referentially opaque, the user of that concept may not be able to specify its exact extension. If all of our concepts were referentially transparent, we would be able to specify the precise extensions of anything we were considering. As a result, we could not have two different concepts which, unbeknownst to us, happen to refer to the same object or class of objects. Since we would be aware of the extensions of the two concepts, we would be aware that their extensions were the same. Yet, this phenomenon is fairly typical. For example, imagine that I have a concept of table salt as a white grainy substance with a characteristic taste that can enhance the taste of certain foods when used in moderation. I have learned this concept by ostension. I also have a concept of sodium chloride as some sort of chemical compound which is poisonous. We can further imagine that I am out of table salt and my friend has brought me a canister of sodium chloride to fill my salt shaker. If I were to claim that the stuff in the canister is poisonous, she would respond that it is just table salt. If Moore's formulation of the Open Question Argument were sound, I could reply that the contents of the canister couldn't be table salt since for any sample of sodium chloride I might consider, it is intelligible for me to question whether or not that sample of sodium chloride really is table salt. But the fact that this is an open question for me does not show that the sodium chloride in the canister before me is not table

salt. Sodium chloride is table salt. By the same token, the mere fact that it is possible to ask of any proposed analysis of 'good', whether it is itself good, does not show that the proposed analysis does not isolate the property of goodness. For this reason, there is a serious problem with Moore's formulation of the Open Question Argument.

1.6: Can Moore's Argument Be Salvaged?

Needless to say, the semantic theory which undergirds Moore's formulation of the Open Question Argument has long since been rejected by the philosophical community. Since that time a number of different theories of meaning gained prominence, most of which equated the meaning of a term with some version of a description or formula which enables us to isolate all and only the referents of that term. This trend has been supplanted by the innovative externalist theories of reference that were pioneered by Kripke and Putnam in the 1960s. These theories have had a profound effect on contemporary philosophy of language, epistemology, and metaphysics. The Cornell Moral Realists contend that this new perspective on the nature of meaning and reference provides us with sufficient grounds for rejecting the Open Question Argument as a hopelessly obsolete approach to moral theorizing which requires no further consideration. According to them, Moore's flawed conception of language and meaning resulted in his mistakenly thinking that he had shown us something significant about the nature of goodness when, in fact, he had

done nothing of the kind. There is really nothing more to the Open Question Argument than that. Once we understand this, **Principia Ethica** can be safely regarded as a mere period piece which is only of historical interest. We can now go on to formulate an updated naturalistic moral semantics à la Kripke. Other moral theorists are not so quick to dismiss Moore's approach, however. According to Darwell, Gibbard, and Railton:

To grant Moore all of the resources he deploys or assumes in his official presentation of the open question argument would suffice to bring the whole enterprise of conceptual analysis to a standstill, and show nothing about Good in particular. ...

Why, then, isn't Moore's argument a mere period piece? However readily we now reject as antiquated his views in semantics and epistemology, it seems impossible to deny that he was on to something.²²

It might be the case that Moore was making a legitimate point even though he was expressing it through a flawed semantics and epistemology. If so, then perhaps the point can be reformulated in a way that does not rely on Moore's questionable semantics and epistemology. In what follows, I will attempt to demonstrate that this is indeed the case. I will begin by arguing that the application of Kripke-Putnam semantics to moral terminology will not put moral realism on a firmer footing. This is because contemporary moral realism is open to criticism by a new form of the Open Question Argument that is not based on Moore's dated theories of semantics and epistemology. Treating moral terms like natural kind terms commits the moral realist to a position which is

²²Darwell, Gibbard, Railton, 1992, pp.115,116.

strongly anti-intuitive in a significant range of scenarios. Our intuitions concerning this range of scenarios can be much more adequately accounted for by adopting a noncognitivist characterization of moral language use. Thus, noncognitivism is preferable over moral realism.

1.7: Kripke's Theory Of Reference

In order to accomplish this task, it seems best to take some time to elaborate on the nature of the new theory of reference which supposedly renders the Open Question Argument obsolete. We can begin by noticing that, as human beings, we have the ability to imagine what the world would be like if certain aspects of it were different. That is to say, we can conceive of different possible worlds besides the actual one. We can also notice that various ways of referring to the objects and types of objects we might encounter in the actual world are available to us. For example, we can call them by name, refer to some of their salient qualities, provide a verbal specification of their location, or even point to them. Some of these ways will designate the same objects or types of objects in all possible worlds; others will not. Kripke introduces the following terminology in this regard.

Let's call something a **rigid designator** if in every possible world it designates the same object, a **nonrigid** or **accidental designator** if that is not the case.²³

One of the key themes of Kripke's theory of reference is that the notion of objects having accidental and essential

²³Kripke, 1972, p.48.

properties is directly related to the notion of rigid designation. An accidental property is a property that an object does not necessarily have to have in order to be an object of the type that it is. On the other hand, an essential property is a property that an object must possess in order to be of the particular type that it is. An important point to notice is that the set of criteria we use in practice to establish whether or not an object is of a particular type need not be the same as the set of criteria which must be satisfied in order for that object to be of that type. The concept that we have of some item in the world might turn out to be a collection of essential and accidental characteristics which we have come to commonly associate with that item. Often, it is easier to learn to identify items in terms of their easily detectable accidental characteristics instead of in terms of their essential characteristics, which may be much more difficult to detect. Kripke applies this point to individual objects when he discusses the cases of Aristotle and Hitler.

To me Aristotle's most important properties consist in his philosophical work, and Hitler's in his murderous political role; both, as I have said, might have lacked these properties altogether. ... **Important** properties of an object need not be essential, unless 'importance' is used as a synonym for essence; and an object could have had properties very different from its most striking actual properties, or from the properties we use to identify it.²⁴

²⁴Kripke, 1972, p.77.

These cases clearly illustrate that the properties we take to be the identifying characteristics of particular individuals need not be the properties which are essential for being those individuals. Instead, the properties of an individual which we take to be the most important may be entirely accidental. Kripke uses this observation as the basis for an alternative account of reference. The definite description view that he is criticizing relies on the set of

properties which somehow qualitatively uniquely pick out an object and determine our reference in that manner.²⁵

Kripke's view, however, focuses on the set of historical conditions in the actual world which have preceded the employment of some particular name by some particular linguistic community. Reference is determined by focusing on the set of conditions in the world which have, in some sense, **caused** an expression to be transmitted throughout a certain linguistic community. In the case of referring to a particular person:

it's in virtue of our connection with other speakers in the community, going back to the referent himself, that we refer to a certain man.²⁶

The referent itself, the actual object in the real world, comes to be connected with a particular name through an informal sort of ceremony which must be an actual event. The

²⁵Kripke, 1972, p.94.

²⁶Kripke, 1972, p.94.

referent must be connected to the name as a result of what **did** happen; not what **might** have happened. In Kripke's words:

An initial 'baptism' takes place. Here the object may be named by ostension, or the reference of the name may be fixed by a description. When the name is 'passed from link to link', the receiver of the name must, I think, intend when he learns it to use it with the same reference as the man from whom he heard it.²⁷

Once this connection has been made, certain statements which contain that name become **necessarily true**. Since necessity is a metaphysical notion rather than a semantic one, this necessity accrues from the nature of the referent which is rigidly designated by that name. Statements which predicate any of the referent's essential properties of the referent are necessarily true.

These remarks about the way that the referent for a name is determined also apply to the way that a class of referents for a natural kind term is determined. As Kripke points out:

my argument implicitly concludes that certain general terms, those for natural kinds, have a greater kinship with proper names than is generally realized.²⁸

He uses the example of gold to illustrate that the set of historical conditions in the actual world which have preceded the employment of some natural kind term by some particular linguistic community are what determine the reference of that term.

we use 'gold' as a term for a certain **kind** of thing. Others have discovered this kind of thing and we have

²⁷Kripke, 1972, p.96.

²⁸Kripke, p.134.

heard of it. We thus as part of a community of speakers have a certain connection between ourselves and a certain kind of thing. The kind of thing is **thought** to have certain identifying marks. Some of these marks may not really be true of gold. We might discover that we are wrong about them. Further, there might be a substance which has all the identifying marks we commonly attributed to gold and used to identify it in the first place, but which is not the same kind of thing, which is not the same substance.²⁹

As in the naming case, statements which predicate any of the reference class's essential properties of the reference class or members of the reference class are **necessarily true**. The essential properties which make the reference class the kind that it is are determined by the actual nature of the set of objects in the world which precipitate the natural kind term predications made by the actual linguistic community. We can see this by considering counterfactual scenarios where the nature of a set of objects has been altered. Suppose, for example, that all the cats in the world turned out to be fur-covered reptiles. In such a case:

the inclination is to say, not that there turned out to be no cats, but that cats have turned out not to be animals as we originally supposed. The original concept of cat is : **that kind of thing**, where the kind of thing can be picked out by paradigmatic instances.³⁰

Once it has been established that actual cats are in fact animals and not reptiles, however, it is necessary that all cats are animals. Any particular cat is necessarily an animal;

²⁹Kripke, 1972, pp.118,119.

³⁰Kripke, 1972, p.122.

though this is something which could not have been known a priori.

This characterization of reference has important implications for the way that possible worlds have to be evaluated. Given that a substance like gold has certain essential properties in the actual world, such as having the atomic number 79, any substance in any possible world which does not have the atomic number 79 cannot possibly be gold. If we do refer to some substance in some possible world which does not have the atomic number 79, then it is necessary that the substance we are referring to is not gold. Kripke invites us to consider the following scenario:

Suppose that all the areas which actually contain gold now, contained pyrites instead, or some other substance which counterfeited the superficial properties of gold but lacked its atomic structure. Would we say, of this counterfactual situation, that in that situation gold would not even have been an element (because pyrites is not an element)? It seems to me that we would not. We would instead describe this as a situation in which a substance, say iron pyrites, which is not gold, would have been found in the very mountains which actually contain gold and would have had the very properties by which we commonly identify gold. But it would not be gold; it would be something else. One should **not** say that it would still be gold in this possible world, though gold would then lack the atomic number 79. It would be some other stuff, some other substance.³¹

This illustration shows that the important criterion which determines whether two substances are of the same type is the way that the essential properties of the substance in the actual world compare to the substance being considered in the

³¹Kripke, 1972, p.124.

possible world. The superficial properties which people might use to identify instances of a substance are not what determine its kind. They are simply the means by which we come to fix the reference class which constitutes that particular kind. When we make predications using the natural kind term in question we mean to refer to the reference class so fixed. We don't intend to refer to the class of all objects which might happen to possess the superficial identifying characteristics that have been used to fix the natural kind. We intend to refer to the class of objects which **are** fixed as a result of our initial use of the superficial identifying characteristics. We need to look at the nature of the class so fixed rather than the superficial identifying characteristics in order to determine which necessary truths pertain to the reference class. As Kripke points out:

one should bear in mind the contrast between the a priori but perhaps contingent properties carried with a term, given by the way its reference was fixed, and the analytic (and hence necessary) properties a term may carry, given by its meaning. For species, as for proper names, the way the reference of a term is fixed should not be regarded as a synonym for the term.³²

Whether or not the initial superficial identifying characteristics by which we fix a reference class are part of the essential nature of that reference class is a purely contingent matter. Thus, we cannot know a priori whether any of the identifying characteristics that we use to fix a

³²Kripke, 1972, p.135.

reference class are essential or accidental features of that class.

1.8: Putnam's Twin Worlds

Putnam has provided a parallel argument for the claim that the meaning of the terms we use cannot be determined simply by analysing the psychological profiles of typical users of such terms. He invites us to consider a situation where two planets are completely identical except that wherever the clear liquid which is constituted by H^2O is found on the first planet, a liquid which is similar in all of its surface characteristics but is constituted of the complex substance XYZ is found on the second planet. We can call these two worlds 'Earth' and 'Twin-Earth'. In each of the two worlds, people refer to the clear liquid which is typically found in rivers and lakes as 'water'. In so referring, they each have the same concept or intension of water as the clear liquid that we see in rivers and lakes, drink out of faucets, bathe in, etc. As a result, when they think of water, they are in the same psychological state. Putnam notes that although the term 'water' has the same intension for speakers on both worlds, it clearly does not have the same extension, since, by stipulation, 'water' refers to a different substance in each of the respective worlds. Thus, the moral we can draw is that the intension of a term does not determine its extension. We have to look beyond what a speaker intends by her use of a term to determine what her uses of that term actually mean. In

Putnam's famous words: "Cut the pie any way you like 'meanings' just ain't in the head."³³

An important corollary of this view of meaning is that people are only in a position to **discover** that the two liquids are not of the same type once they have gained a certain degree of understanding of the underlying nature of the substances around them. Yet, different members of a linguistic community will attain different degrees of understanding of the underlying nature of these substances. This leads to the phenomenon which Putnam describes as 'the division of linguistic labor'.

When a term is subject to the division of linguistic labor, the "average" speaker who acquires it does not acquire anything that fixes its extension. In particular, his individual psychological state **certainly** does not fix its extension; it is only the sociolinguistic state of the collective linguistic body to which the speaker belongs that fixes the extension.³⁴

The point of Putnam's metaphor is that the concepts which typical speakers have for many natural kind terms are quite distinct from what would be necessary to be able to determine the essential nature of the class of referents which correspond to such natural kind terms. The concepts people have of many natural kinds are generally operational procedures which enable them to identify instances of such kinds quickly and efficiently. Their practical interest is simply to refer to the kind in question; not to specify its

³³Putnam, 1973, p.311.

³⁴Putnam, 1973, p.312.

essential nature. Thus, the concept most of us have of water is a clear liquid which comes from rivers and lakes and is good for drinking and bathing. This concept enables us to identify instances of water and speak about it in the contexts in which we generally find ourselves. But the Earth/Twin-Earth scenario shows that we don't take our concept of water to explicate the essential nature of water. As Putnam notes:

If I agree that a liquid with the superficial properties of "water" but a different microstructure **isn't really water**, then my ways of recognizing water cannot be regarded as an analytical specification of what **it is to be** water. Rather, the operational definition, like the ostensive one, is simply a way of putting out a standard - pointing out the stuff **in the actual world** such that, for **x** to be water, in **any** world, is for **x** to bear the relation **same_l** to the **normal** members of the class of **local** entities that satisfy the operational definition. "Water" on Twin Earth is not water, even if it satisfies the operational definition, because it doesn't bear the **same_l** to the **local** stuff that satisfies the operational definition, and the local stuff that satisfies the operational definition but has a microstructure different from the rest of the local stuff that satisfies the operational definition isn't water either, because it doesn't bear **same_l** to the **normal** examples of the local "water".³⁵

Thus, we can see that the predications we make using natural kind terms actually have an indexical element to them. Part of the meaning of such terms is determined by the actual nature of the reference class which happens to be picked out by our operational notions. Thus, we have to go beyond our current sociolinguistic agreement to determine the meaning of our natural kind terms. We have to consult the way that the world

³⁵Putnam, 1973, pp.313,314.

actually turns out to be. But, as Putnam points out, speakers may be unaware that such facts pertain to the reference class.

The extension of our terms depends on the actual nature of the particular things which serve as paradigms, and this actual nature is not, in general, fully known to the speaker. Traditional semantic theory leaves out two contributions to the determination of reference - the contribution of society and the contribution of the real world; a better semantic theory must encompass both.³⁶

As a result, there may be facts about the reference class which a competent speaker may be entirely unaware of, since he only thinks about the reference class in terms of its operational definition.

1.9: Moore's Open Question Argument Rendered Obsolete

An important unintended result of this new semantics is that it can be used to provide a clear explication of precisely what is wrong with Moore's formulation of the Open Question Argument. Recall that Moore believed the meaning of a term is determined entirely by the concept or intention we have in mind when we make predications using that term. According to Moore, we can see that no natural property can provide an analysis of what it is to be good because we intend something different when we say that an object is good than we do when we say that an object possesses that natural property. Putnam characterizes one of the crucial theoretical suppositions underlying Moore's Open Question Argument in the following manner.

³⁶Putnam, 1973, p.315.

Moore's argument turns on assumptions that I and many other philosophers of language would reject today, however. First of all, he implicitly denied that there could be such a thing as **synthetic identity of properties**.³⁷

Moore thought that the only way of settling the issue of whether or not two properties are identical is by comparing the intentions that a person has when he predicates the first property of something with the intentions that he has when he predicates the second property of that thing. If the person intended the same thing by the first predication as he did by the second, then he was referring to the same property in both predications. On the other hand, if the person intended something different by the first predication from what he did by the second, then the properties he was referring to in each of the respective predications must have been different. Whenever it is intelligible for him to question whether or not the making of one predication entails the making of the other, then the speaker did not mean the same thing by each of the two predications. This is sufficient to show that the properties corresponding to each of the respective predications are not one and the same.

The crucial problem with this position is that it makes an implausible connection between what goes on in a person's head and what is the case in the external world. According to Putnam: "Moore conflated **properties** and **concepts**."³⁸ He didn't

³⁷Putnam, 1981, p.206.

³⁸Putnam, 1981, p.207.

see that it is possible for a person to have two different sets of criteria that specify the conditions under which it is appropriate to predicate some property of an item (and, thus, two distinct concepts which turn out to track the same property in the world). That the two distinct concepts do, in fact, track the same property is something we discover by empirical investigation rather than through introspective analysis. To discover that two distinct concepts turn out to track the same property is to discover something about the nature of the world; not merely to discover something about the nature of the two concepts in question. It is to discover the **synthetic truth** that the properties referred to by the two concepts are, in fact, one and the same property. Thus, the mere fact that it is intelligible to question whether the presence of the property corresponding to the making of the first predication entails the presence of the property corresponding to the making of the second predication does not preclude the possibility that the two properties are actually one and the same. Thus, it does not show that the meaning of a term like 'good' cannot be cashed out entirely in terms of some natural property or cluster of natural properties. In this way, the externalist theory of reference inspired by Kripke and Putnam can be used to articulate precisely what is wrong with Moore's original formulation of the Open Question Argument.

1.10: An Updated Semantics For Moral Terminology

Having taken the sting out of Moore's influential argument, The Cornell Moral Realists wholeheartedly endorse the application of this updated theory of reference to the moral realm. There are two key aspects of the contemporary moral realism they espouse. The first aspect concerns the nature of the reference relation between moral terms and types of natural states of affairs. According to The Cornell Moral Realists, as with any other natural kind term, the meaning of a moral expression is determined by the states of affairs which actually occasion its use. For example, the meaning of 'good' is constituted by the set of conditions which generally precipitate attributions of goodness by competent speakers. We can think of these states of affairs as possessing a common property or as being of a common type. The relationship between properties and types is fairly straightforward. A property can be taken as the criterion which determines whether or not objects are of a particular type. If an object possesses the requisite property, then it is of that type. Thus, objects are of the type green, if and only if, they possess the property of greenness. Persons, actions, and states of affairs are of the type good, if and only if, they possess the property of goodness.

The question the Cornell Moral Realist is considering is whether or not some type of natural state of affairs constitutes those states of affairs which are the good ones.

If we designate some type of natural state of affairs when we use the word 'good', then we are using that word to refer to a natural kind which is out there to be so designated. As is the case with other natural kind terms, once this use has been established, the only states of affairs which can count as good are those which share the same essential qualities as the type of states of affairs which has established the meaning of 'good' in the actual world. It is not possible for any natural states of affairs which do not possess these essential qualities to be good. 'Good' rigidly designates a particular property which is only instantiated in certain states of affairs. Once this use has become established, certain truths about which states of affairs are good and which are not will have become necessary. Such truths are a posteriori necessities since they did not become necessary until after the historical events that established the particular meaning which 'good' now has had occurred.

In order to elucidate the nature of the meaning that 'good' has come to acquire, Richard Boyd has developed a sophisticated version of the causal theory of reference called 'homeostatic property-cluster definitions'.

According to various property-cluster or criterial attribute theories, some terms have definitions which are provided for by a collection of properties such that the possession of an adequate number of these properties is sufficient for falling within the extension of the term.³⁹

³⁹Boyd, 1988, p.196.

Boyd goes on to maintain that the extensions of our moral terms constitute natural kinds which are most suitably characterized as homeostatic property clusters. We have an instance of a natural kind when there is an objective tendency for certain properties to cluster together. In his words

Their co-occurrence is not, at least typically, a statistical artifact, but rather the result of what may be metaphorically (sometimes literally) described as a sort of homeostasis.⁴⁰

That these properties tend to cluster together as they do is not an a priori fact, however; it is simply the way that we find the world to be. In the non-moral case

'Water = H₂O' is a non-analytic necessary truth that expresses the real, underlying essence of water and thus provides a (synthetically true) definition of 'water'.⁴¹

In the moral case, 'Goodness = a certain homeostatic cluster of properties' is also a non-analytic necessary truth that expresses the real, underlying essence of goodness and thus provides a (synthetically true) definition of 'goodness'.

The second key aspect of this new version of naturalistic moral realism concerns the nature of the relationship between moral properties and straightforward physical properties. In his 'argument from queerness' Mackie claimed that any putative connection between natural and moral properties would remain intractably mysterious. As he says:

⁴⁰Boyd, 1988, p.197.

⁴¹Horgan and Timmons, 1990, p.451.

Another way of bringing out this queerness is to ask, about anything that is supposed to have some objective moral quality, how this is linked with its natural features. What is the connection between the natural fact that an action is a piece of deliberate cruelty - say, causing pain just for fun - and the moral fact that it is wrong? It cannot be an entailment, a logical or semantic necessity. Yet it is not merely that the two features occur together. The wrongness must somehow be 'consequential' or 'supervenient'; it is wrong because it is a piece of deliberate cruelty. But just what **in the world** is signified by this 'because'?⁴²

Recent developments in the philosophy of science and the philosophy of mind open the way for an answer to this question by making the notion of supervenient or functional properties respectable within a naturalistic framework.

According to David Brink,

Supervenience is a relation of causal constitution or dependence. There is nothing strange and certainly nothing unique about the supervenience of moral properties on physical properties.⁴³

These second order properties are real but **not identical** to the physical base properties upon which they supervene.

The realist might claim that moral properties are those which bear on the maintenance and flourishing of human organisms. ... The physical states which contribute to or interfere with the satisfaction of these needs, wants, and capacities are the physical states upon which, on this functional theory, moral properties ultimately supervene.⁴⁴

It seems to me that both aspects of The Cornell Moral Realists' naturalistic version of moral realism are problematic. I will begin by criticizing the second aspect.

⁴²Mackie, 1977, p.40.

⁴³Brink, 1984, p.120.

⁴⁴Brink, 1984, pp.121,122.

1.11: Supervenience Accounts Of Moral Properties

As Brink acknowledges, a problem with this type of approach is that an implicit normative judgement has already been made in deciding what is to count as human well-being and flourishing. But he claims that the fact that such judgements are theory-dependent in this way does not present a **special** problem for the moral realist. This reply is not satisfactory, however, since the problem is not simply that judgements concerning well-being or flourishing are theory-dependent but that they depend on a **normative** theory at point where we are still trying to establish the conditions for a purely naturalistic normativity. Further, even if we suppose that there is a best theory which delineates the manner in which moral properties supervene upon natural properties in the individual case, there is no reason to think it can be automatically transposed to the social case. Even people who are in complete agreement about which conditions are conducive to the realization of a good life for an individual sharply disagree about what the best way of apportioning social opportunities is when conditions of scarcity make it impossible for all parties to realize their optimal potential. We can't assume in advance that there is a **best** way of dealing with this problem.

One might attempt to deal with this problem by appealing to moral theory. According to Brink:

Which moral properties strongly supervene on which natural properties is determined differently by different

moral theories, just as, say, which economic properties strongly supervene on which social and psychological properties is determined differently by different economic theories.⁴⁵

The fact that the supervenience relation is theory-dependent is not taken to be a problem here because theory dependent relations in the scientific world are common. In order to arrive at the genuine moral properties, however, we will not want to rely on just any moral theory. We will want to ensure that we are relying on the best moral theory possible since it will play a role in determining which moral properties there are and what they are like. In Brink's words:

Determination of which account of supervenience is best will depend upon determination of which moral theory provides the best account of all our beliefs, both moral and non-moral.⁴⁶

In this approach, the mechanism of wide reflective equilibrium (WRE) enables us to **discover** the best moral theory which in turn enables us to **discover** the correct account of the manner in which moral properties supervene on physical properties. What WRE does is enable us to zero in on the natural kinds which we have been loosely referring to with our moral terminology all along. There is a body of external, objective, synthetic moral data which can be discovered through WRE. As in science, we give prima facie epistemic weight to our initial beliefs and proceed to subject them to theoretical scrutiny and testing. Eventually, we will arrive at a reliable

⁴⁵Brink, 1989, p.175.

⁴⁶Brink, 1984, p.121.

set of moral beliefs which constitute the best moral theory. In the end, Brink seems to glibly assume that there is a **best** moral theory that we can ultimately rely on to specify the character of the complex natural base properties upon which moral properties supervene.

Darwell, Gibbard, and Railton argue that this sanguine hope of arriving at a best moral theory is not without its problems. We cannot simply assume that the best moral theory can be arrived at in roughly the same manner as the best scientific theory. We can appeal to the success of a scientific theory in order to bolster its credibility, but there does not seem to be an analogue to scientific success in the moral realm. As they point out:

critics will reply by citing many differences in the actual **accomplishments** of theory development in science as opposed to ethics. ... the lack of consensus and accomplishment in ethics counts against what would otherwise be a reasonable willingness to assign **prima facie** epistemic weight to moral beliefs. There would of course be little left of ethics, or ethical inquiry, as a cognitive domain were no initial weight given to going moral beliefs; but critics might not find this skeptical result wholly uncongenial.

An effective reply to such critics would involve giving an explanation of moral controversy that competes favorably with the skeptical alternative and, more importantly, would involve developing nonreductionist ethical theory itself, showing it to have (or to make possible) some worthwhile theoretical accomplishments.⁴⁷

Such a defense is still forthcoming. Sturgeon offers a kind of defense when he attempts to parallel scientific and moral theory with the following example.

⁴⁷Darwell, Gibbard, Railton, 1992, p.173.

Candidate moral principles - for example, that an action is wrong just in case there is something else the agent could have done that would have produced a greater net balance of pleasure over pain - lack empirical implications when considered in isolation. But it is easy to derive empirical consequences from them, and thus to test them against experience, if we allow ourselves, as we do in the scientific case, to rely on a background of other assumptions of comparable status. Thus, if we conjoin the act-utilitarian principle I just cited with the further view, also untestable in isolation, that it is always wrong deliberately to kill a human being, we can deduce from these two premises together the consequence that deliberately killing a human being always produces a lesser balance of pleasure over pain than some available alternative act; and this claim is one any positivist would have conceded we know, in principle at least, how to test.⁴⁸

Such an example can hardly be regarded as a resounding theoretical accomplishment that provides cogent support for the theory of act-utilitarianism. The problem with Sturgeon's example is not that such a testing procedure would provide no support at all for act-utilitarianism. The problem is that analogous testing procedures could be used to provide the same kind of support for any number of competing normative perspectives. We would still be left with the problem of determining which of these competing normative perspectives is the **best** moral theory. Those who wish to speak of the best moral theory still have the onus on them to demonstrate that this can be done intelligibly. The Cornell Moral Realists have yet to satisfy the dictates of this onus. Since their supervenience account of the nature of moral properties

⁴⁸Sturgeon, 1984, p.51.

requires that there is a best moral theory, the account remains dubious.⁴⁹

Having noted this problem, we can summarize the semantic position of The Cornell Moral Realists. According to their updated causal theory of reference for moral terminology, whether or not the average person is fully aware of this fact, moral properties are what we are actually referring to when we employ moral terminology. For example, what makes a statement like 'Acts of senseless cruelty are bad' true, is that the best moral theory specifies that the moral property of badness supervenes on the natural class of actions which includes acts of senseless cruelty. The predicate 'bad' rigidly designates a certain class of natural states of affairs. If I say 'What she did is bad', this statement will be true iff her action resulted in an instantiation of that kind of natural state of affairs. It is crucial to notice that our best moral theory is not a part of these truth conditions; it merely plays the role of specifying what the classes of states of affairs which constitute the moral properties actually consist of. For any moral statement describing a particular state of affairs, it is still the instantiation of the requisite cluster of natural properties which **makes** that moral statement true; our best moral theory merely determines **what** makes that moral statement true. Once the homeostatic cluster of natural properties that

⁴⁹In Chapter Four I will argue that as a result of the very nature of the problem they **can't** satisfy this onus.

constitutes the reference class for a moral term has been rigidly designated, any other cluster of natural properties will be of a different kind from the one which has already been rigidly designated. When people use the moral term, they are referring to **that** moral property, opaque as such a reference may be to some of them. Users of moral terminology need not have any attitude towards or detailed awareness of our best moral theory even though it demarcates the parameters of the reference classes corresponding to the moral terminology they employ. A typical moral language user may have no clearer conception of the nature of this theory than the average speaker of English has of the theory which demarcates the reference class of the group of related ailments we rigidly designate by the term 'cancer'.

1.12: Dispositional Accounts Of Moral Properties

Before criticizing the semantic theory of The Cornell Moral Realists in detail, I will pause to explain why a dispositional account of the nature of moral properties also seems to be an implausible option for the moral realist. In a dispositional account of moral realism "ethical facts and properties must be understood in terms of our dispositions to respond".⁵⁰ The problem with this kind of approach is that it makes our dispositions to respond a crucial part of the truth conditions for moral statements. As such, it no longer allows our moral terms to rigidly designate particular kinds of

⁵⁰Brower, 1993, p.221.

natural states of affairs. As a result, it is unable to provide a realist construal of the meaning of moral terms.

Kripke makes some interesting comments about the nature of heat which can be used to explain why dispositional accounts of moral properties are problematic. In the history of our usage of the term 'heat', we initially identified heat as that property of objects which is able to produce a particular sensation in us which we can call S. We have since discovered that this S-producing property of objects corresponds to the rate of movement of their constituent molecules. That is to say, we have discovered that heat is mean molecular kinetic energy. Heat is a natural kind whose presence we are able to detect by the sensation S that it produces in us. Kripke notes that:

Someone can be in the same epistemic situation as he would be if there were heat, even in the absence of heat, simply by feeling the sensation of heat; and even in the presence of heat, he can have the same evidence as he would in the absence of heat simply by lacking the sensation S. ... In the case of the identity of heat with molecular motion the important consideration was that although 'heat' is a rigid designator, the reference of that designator was determined by an accidental property of the referent, namely the property of producing in us the sensation S.⁵¹

If increased mean molecular motion failed to produce sensation S in another group of beings, we would not say that, in that situation, mean molecular kinetic energy is not heat. Rather, we would say that heat does not produce in them the sensation by which we can easily detect its presence. Heat is a

⁵¹Kripke, 1972, p.152.

phenomenon in the actual world whose nature, once the reference of 'heat' has been fixed, is determined quite independently of the sensations it causes and the responses it evokes in various groups of beings. If goodness is also a natural property, then, once its reference has been fixed, its nature will also be determined quite independently of the sensations it causes and the responses it evokes in various groups of beings. Neither goodness nor heat can be defined in terms of a certain kind of sensation which it causes or a certain kind of response which it evokes if a causal theory of reference is adopted. The reference class which corresponds to the natural kind term in question must be the natural state of affairs which has been rigidly designated by that natural kind term. If the class of natural states of affairs A evokes the same response in us as the class of natural states of affairs B evokes in another group of people, and we each rigidly designate our respective response-evoking class of states of affairs by 'good', then we each mean something different by 'good'. Continuity of reference cannot be established by pointing to the similar response that each of the two classes of states of affairs evokes in its respective group. Thus, dispositional accounts of moral properties are not open to the naturalistic moral realist for the same reasons that dispositional accounts of natural properties like heat are not open to any naturalistic realist account.

1.13: An Initial Problem For Updated Moral Semantics

At this point, I will move on to criticize the first key aspect of The Cornell Moral Realists' contemporary version of naturalistic moral realism. This is the application of the Kripke-Putnam theory of reference to moral terminology. Before considering direct criticisms of this new semantics for moral terminology, we can begin by noting an important difference between the putative natural kind **good** and other more straightforward natural kinds. Consider the following statement which contains the natural kind term 'water':

(1*) 'This liquid is not water even though it is H^2O .'

According to the Kripke-Putnam semantics, if it is the case that water is comprised of H^2O , then (1*) is necessarily false. If one is committed to the view that water is comprised of H^2O , one is thereby committed to the view that (1*) is necessarily false. According to our best scientific theories, our use of the word 'water' rigidly designates a particular kind of liquid which is comprised by the particular molecular configuration which is rigidly designated by ' H^2O '. That the word 'water' has come to rigidly designate this particular liquid is a matter of contingent fact. The word 'water' could have been used to rigidly designate another kind of substance. Now that the reference of 'water' has been fixed, however, the only way we can be referring to H^2O is if we are at the same time referring to water. To refer to a liquid which is not water is, by that very token, to refer to a liquid which is

not H²O. (1*) is a necessary falsehood which could not have been known to be such a priori. Thus, any scientifically literate person who accepts the Kripke-Putnam semantics will take (1*) to be not just false, but necessarily false.

It is significant to notice that we don't see a parallel position arising in the moral case. If 'good' really is a natural kind term just like 'water', then the utilitarian should be committed to the position that

(1) 'this action is not good even though it is conducive to maximizing total utility'⁵² is not just false, but **necessarily false**. According to the Kripke-Putnam semantics, if it is the case that goodness is constituted by maximal total utility, then (1) has to be necessarily false. If one is committed to the view that goodness is constituted by conduciveness to maximizing total utility, one is thereby committed to the view that (1) is necessarily false. According to a utilitarian version of moral realism, our use of the word 'good' rigidly designates those states of affairs which are also rigidly designated by 'conducive to maximizing total utility'. Now that the reference of 'good' has been fixed, the only way we can be referring to an action which is conducive to maximizing total utility is if, at the same time, we are referring to an action which is good. To refer to an action which is not good is, by

⁵²Putnam, 1981, p.206.

that very token, to refer to a action which is not conducive to maximizing total utility. (1) is a necessary falsehood which could not have been known to be such a priori. Thus, any committed utilitarian has to take (1) as not just false, but necessarily false. When people speak of actions which, say, conform to the categorical imperative, they cannot be speaking of actions which are good any more than a person who is speaking of liquids which are comprised of XYZ can be held to be speaking of water. The utilitarian moral realist must count them as speaking of actions which are schmood, where 'schmood' rigidly designates a class of actions which is different from that class of actions which the utilitarian moral realist rigidly designates by 'good'.

Suppose we fix the reference of a natural kind in the actual world by describing it in terms of one or more of its contingent, but easily identifiable properties. For example, we might think of crows as the black birds which are often heard to make distinctively annoying 'caw' sounds. Then, we can speak of instances of that natural kind in various possible worlds which lack the reference-fixing contingent property or properties, but are easily identified by people in that world through other reference-fixing contingent properties. For example, we can imagine a possible world where people think of crows as the black birds which are often heard to make pleasant, harmonious chirping sounds. When we do so, we are referring to instances of a kind in the possible world

which is the same as the kind in the actual world whose reference has already been fixed. In such a case, we do not mean the same thing by the natural kind term as we do by the description we use to fix the reference of that kind in the actual world. Here, the natural kind term and the description we use to fix the reference of the natural kind term are not synonymous. As a result, it makes sense to ask whether or not an instance of the natural kind so fixed actually satisfies the description. For any particular crow we might imagine, it is an open question whether or not that crow is black and is prone to make distinctively annoying 'caw' sounds.

On the other hand, suppose we fix the reference of a natural kind in the actual world by describing it in terms of one or more of its **essential** properties. In this case, we cannot speak of instances of that natural kind in various possible worlds which happen to lack the reference-fixing essential property or properties. If the entities in question did not have one or more of the essential properties of that natural kind, then they could not be instances of **that** kind. The important thing to notice is that once we have **discovered** that certain properties are essential properties of certain natural kinds, certain natural kind terms **become** synonymous with certain descriptions. For example, at one time (1*) wasn't self-contradictory, in the sense of being analytically false, because we didn't know that being H^2O is an essential property of being water. But things have changed. Most people

are well aware that water consists of H^2O . Now that we know that being H^2O is an essential property of being water, (1*) is self-contradictory. Questions concerning whether or not an instance of a natural kind satisfies the description used to fix the reference of that natural kind can only be open questions when the reference-fixing description is comprised of contingent properties of that natural kind or essential properties of that natural kind which are not known to be essential. Open questions are not possible when we are considering whether or not instances of a natural kind satisfy descriptions which consist of **known essential properties** of that kind. Thus, our knowledge of the ultimate constituents of natural kinds has an effect on which questions we will treat as open. My claim is that this is so in the non-moral case but not in the moral case.

Consider the utilitarian moral realist again. If he believes that certain actions are good in virtue of their being conducive to total maximal utility, then surely he believes that their being conducive to total maximal utility is an essential rather than accidental property of such good actions. He could understand how (1) could be an open question for those who have not yet realized the fact that our uses of 'good' track the complex natural property of being conducive to total maximal utility, but for him it would have to be necessarily false. If he doesn't take (1) as necessarily false, he needs to explain why he doesn't when he takes (1*)

to be necessarily false. Why is he treating these two natural kind terms differently?

1.14: Moral Twin-Earth

The difficulty that has been alluded to in the previous section is brought out more directly by Horgan and Timmons in a recent article entitled "New Wave Moral Realism Meets Moral Twin Earth."⁵³ They invite us to consider a moral version of Putnam's Twin-Earth thought experiment. In this version, Earth and Moral Twin-Earth are two planets which are exactly alike in all respects except that the functional properties which regulate the use of moral words like 'good' and 'right' turn out to be different on the respective planets. On Earth, the normative theory which best captures the use of moral predicates is a consequentialist one which we can call T^c. On Moral Twin-Earth, the normative theory which best captures the use of moral predicates is a deontological one which we can call T^d. The reason for this is that different sets of natural properties occasion the use of moral predicates on each of the two planets. By employing WRE, Earthlings could come

to discover that Earthling uses of moral terms are causally regulated by functional properties whose essence is captured by the consequentialist normative theory T^c.

Twin-Earthlings, on the other hand, would

⁵³Horgan and Timmons, 1990.

discover that their own uses of moral terms are causally regulated by functional properties whose essence is captured by the deontological theory T^d .⁵⁴

Given this setup, what are we to make of the fact that different functional properties which supervene upon different homeostatic clusters of natural properties regulate the use of the same moral words on the two planets? If the meaning of moral terms is adequately characterized by the sophisticated causal theory of reference endorsed by The Cornell Moral Realists, then we should conclude that Earthlings and Moral Twin-Earthlings simply mean different things by utterances which contain the same moral terms in the same way that Earthlings and Twin-Earthlings mean different things by utterances which contain the word 'water'. In the latter case, it doesn't make sense to say that Earthlings and Twin-Earthlings have a deep-seated disagreement about the nature of 'wateriness'. By the same token, it should make no sense to say that Earthlings and Moral Twin-Earthlings have a deep-seated disagreement about the nature of morality. Yet, there is an intuitive pull towards characterizing the Moral Twin-Earth scenario in this way.

Suppose that Earthlings visit Twin Earth (or vice versa), and both groups come to realize that different natural properties causally regulate their respective uses of 'good', 'right', and other moral terms. ... recognition of these differences ought to result in its seeming rather silly, to members of each group, to engage in

⁵⁴Horgan and Timmons, 1990, p.460.

inter-group debate about goodness - about whether it conforms to normative theory T^c or T^d .⁵⁵

Since it doesn't seem silly for members of the two planets to engage in such a debate, we have reason to reject the characterization of moral semantics advocated by The Cornell Moral Realists.

1.15: Disagreements Over The Nature Of Justice

Horgan and Timmons have provided us with an important insight here, but it is an insight that needs to be expanded upon. The principal drawback of their account is that their treatment of the Moral Twin-Earth scenario is entirely negative. That is to say, while they provide grounds for thinking that the naturalistic moral realism espoused by the Cornell camp is **not** an adequate characterization of the nature of moral language use, they do not make any headway towards providing a positive account of what **is** an adequate characterization of the nature of moral language use. In what follows, I will explain how an analogous Moral Twin-Earth scenario can be used to argue **in favor** of a noncognitivist characterization of the nature of moral language use. Thus, I will expand on Horgan and Timmons' initial insight in order to make a positive contribution to the moral realism debate. I will begin by discussing the way that our linguistic intuitions are affected when we consider a Moral Twin-Earth scenario where the word 'just' has a different meaning in each

⁵⁵Horgan and Timmons, 1990, p.460.

of the respective worlds. This will illustrate that the contemporary moral realist faces a general problem here which does not only apply to treating moral goodness as a natural kind. Dissimilar linguistic intuitions will arise whenever **any** moral term is juxtaposed with a straightforward natural kind term in Twin-Earth/Moral Twin-Earth scenarios.

Imagine that, on Earth, people use the word 'just' to track the complex natural property of adhering to a Rawlsian scheme of distributive justice. On Moral Twin-Earth, people use the word 'just' to track the complex natural property of adhering to a Nozickean scheme of distributive justice. We can further suppose that there are a pair of successful self-made entrepreneurs, Ernie and Twin-Ernie, each of which has begun a successful athletic shoe business on his respective planet. We can now imagine representatives from the two worlds discussing whether or not it would be **just** to implement a policy of graduated incremental taxation which would result in a large portion of Ernie's and Twin-Ernie's shoe revenues being redistributed to those who are on the lower end of the income distribution curve on their respective planet. As one would expect, the representatives from Earth would maintain that the policy is just while the representatives from Moral Twin-Earth would maintain that the policy is clearly unjust. If the contemporary moral realist characterization of 'just' as a natural kind term which rigidly designates a complex natural property is correct, then this 'disagreement' is no

different in kind from the 'disagreement' Putnam's original Earthlings and Twin-Earthlings would have over whether a carafe of H²O is water or not. As one would expect, the representatives from Earth would maintain that the carafe contains water while the representatives from Twin-Earth would maintain that the carafe does not contain water. Thus, there is no basis for bona fide disagreement in either of the two scenarios since each of the two parties means something different by the natural kind term they are using, be it 'water' or 'just'. Yet, it does not seem plausible to characterize the two parties in this Moral Twin-Earth scenario as merely speaking at cross-purposes as a result of their each using the word 'just' to rigidly designate a different complex natural property. Rather, it seems much more plausible to characterize them as having a bona fide disagreement over which of the two complex natural properties **deserves** to be referred to as just. Implicit in this disagreement is the understanding that to refer to a complex natural property as 'just' is to confer an honorific status upon that property. It is to maintain that the standard of appraisal which condones the existence of that complex natural property is **justified** in some sense.

We can consider a Legal Twin-Earth scenario to illustrate this point. For the purpose of illustration, we can assume that some form of legal positivism is essentially correct so that a policy will be legal in a particular state whenever the

de facto legal apparatus of that state deems such a policy to be legal. The fact that natural law theorists would not accept this assumption does not affect the ability of the scenario to illustrate the point I wish to make here. Suppose that, on Earth, the Canadian constitution is set up in such a way that a policy of graduated incremental taxation could easily be legally implemented. On Legal Twin-Earth, the Twin-Canadian constitution is set up in such a way that a policy of graduated incremental taxation could never be legally implemented. If representatives from the two worlds disagreed over whether or not the implementation of a policy of graduated incremental taxation in Canada would be **legally just**, it would be most natural to take them as arguing at cross-purposes. Each of the two parties would be using the legal system in the country he calls Canada to determine whether or not the implementation of a policy of graduated incremental taxation is legally just. The two parties would each be using the version of the Canadian constitution that exists on their own world as the standard for specifying which policies or states of affairs are the legally just ones. Yet, in so doing, neither of them would be taking the Canadian legal system on their own world to be a **justified** standard in the sense that the legality of the policies which the standard condones is contingent on that standard's being justified. They would each be of the view that whether or not a policy is legally just in a particular state is purely a function of the

actual character of the legal system that is in place in that state, regardless of whether or not that legal system is taken to be justified in some deeper sense. Thus, there would be no basis for disagreement between the two parties. Once they were both made aware that each of them were using different legal systems to determine the legality of taxation policies, the apparent disagreement would dissolve. This is not so in the moral case, however. Making the two parties aware that they were each using different normative systems to determine which taxation policies are **morally just** would only serve to intensify their disagreement.

The locus of the bona fide moral disagreement between the two parties, then, is over which standard of distributive justice is justified. The two parties have different responses to this issue because they have different conceptions of the nature of justice. This is because they have different ultimate moral perspectives that determine which moral standards are justified. As a result, they are unwilling to accept alternative conceptions of justice as legitimate. This connection of ultimate moral perspectives and justified moral standards to what particular moral language users take to be legitimate moral predication is the crucial difference between Earth/Twin-Earth and Earth/Moral Twin-Earth scenarios. It explains why significant disagreement is possible in the latter range of scenarios but not in the former. The differing ultimate moral perspectives held by the inhabitants of the two

worlds explain their differing conceptions of justified moral standards which in turn explain their differing verdicts on which complex natural properties are to count as genuinely just. At the surface level, the disagreement between the two parties begins over the issue of which complex natural property is appropriately referred to as 'just'. At a deeper level, this is a disagreement over which moral standards are justified. At the deepest level, this is a disagreement over which ultimate moral perspective it is appropriate to endorse.

Nothing like this occurs in Putnam's original Twin-Earth scenario, however. By stipulation, the natural property that each of the two parties refers to in their use of the shared natural kind term just is that natural property which **happens** to occasion their use of that term. In using the natural kind term in the way that they do, neither of the two parties is embracing any ultimate perspective which leads them to endorse a standard which specifies the character of the particular natural property which happens to occasion their use of the natural kind term. As a result, when it is discovered that each of the two parties means something different by the shared natural kind term, since each of their uses of it are occasioned by a different natural property, it becomes clear that there is nothing left for the two parties to be disagreeing about. Neither of the two parties has any **deeper commitments** associated with their using that natural kind term in the way that they do. Thus, they will have no basis for

qualms if it turns out that some other community uses that natural kind term differently. It is only when the two parties have some **deeper commitment about the way that a term is being used** that genuine disagreement in the context of different uses of that term becomes a possibility. The disagreement only becomes genuine in Moral Twin-Earth scenarios because it is only in this range of scenarios that both parties have a shared understanding of the significance of designating a certain complex natural property by a moral term like 'just'.

What we discover, then, is that the employment of moral terminology involves more than the simple rigid designation of certain complex natural properties. It involves the further understanding that the particular complex natural properties so designated have a special status. The persons, actions, and states of affairs which possess that property are the subjects of negative or positive appraisal as a result of their possessing that property. When another group rigidly designates a particular complex natural property with their uses of a moral term which is different from the complex natural property that we designate with our uses of that moral term, they are conferring a negative or positive appraisal on a different set of persons, actions, and states of affairs than the one we are conferring a negative or positive appraisal upon. By our lights, **that** appraisal is reserved for the persons, actions, and states of affairs which possess

another complex natural property, namely, the one that we rigidly designate by our uses of the moral term in question.

At this point, we can reconsider the disagreement between Earthlings and Moral Twin-Earthlings over whether or not a policy of graduated incremental taxation is just. According to my noncognitivist construal of the situation, to call a state of affairs 'just' is to give it a positive appraisal. It is in some sense to express approval of that state of affairs obtaining. To call a state of affairs 'unjust' is to give it a negative appraisal. It is in some sense to express disapproval of that state of affairs obtaining. The class of states of affairs which Earthlings rigidly designate by the moral term 'just' includes those where a policy of graduated incremental taxation is in force while the class of states of affairs which Moral Twin-Earthlings rigidly designate by the term 'just' does not. This is to say that the class of states of affairs which Earthlings approve of includes those which enforce a policy of graduated incremental taxation while the class of states of affairs which Moral Twin-Earthlings approve of does not. The issue between the two parties is over whether or not the policy of graduated incremental taxation belongs in the class of approved states of affairs. It is a debate over which states of affairs ought to be counted as just, and, thus, be approved of, and which shouldn't.

The strategy employed in the preceding discussion can be characterized as a straightforward reductio. We begin by

provisionally assuming that The Cornell Moral Realists are correct in treating moral terms as simply another variety of natural kind terms. If this assumption is correct, then we should expect that moral terms will affect our linguistic intuitions in the same way that other natural kind terms do when we compare the occurrences of moral and straightforward natural kind terms in parallel scenarios. If the usage of moral terms in these scenarios does not affect our linguistic intuitions in the same way that the usage of straightforward natural kind terms does, then we have grounds for claiming that moral terms are not simply another variety of natural kind terms. The fact that our linguistic intuitions are affected differently by the usage of moral terms and the usage of straightforward natural kind terms in these parallel scenarios is a problem for the contemporary moral realist. It is incumbent upon her to be able to explain why our linguistic intuitions are affected differently by moral terms and straightforward natural kind terms if moral terms are really just another variety of natural kind terms. If she can't do this, then, by reductio, we are entitled to conclude that moral terms are not a variety of natural kind terms. Thus, the moral semantics adduced by contemporary moral realists is mistaken and should be rejected. In this way, an updated version of the Open Question Argument can be posed against contemporary moral realists. The problem for the contemporary moral realist is essentially the same as the one posed by

Moore. If naturalistic definitions of moral terms like 'good' or 'just' are accepted, then the possibility of legitimate moral debate is curtailed in advance. We are forced to the anti-intuitive conclusion that if people use terms like 'good' or 'just' to refer to different complex natural properties than we do, then they simply mean something different by these terms than we do, with the result that there is no basis for disagreement. Only a noncognitivist construal of the nature of moral language can account for the possibility of this kind of disagreement. Thus, we have a new argument in favor of noncognitivism.

1.16: Potential Problems With Moral Twin-Earth

There are at least two potential problem areas to be considered in assessing the plausibility of using Moral Twin-Earth scenarios as a means of arguing in favor of noncognitivism. The first potential problem is that Putnam's original Twin-Earth scenario may not be genuinely intelligible. It might be the case that it is not possible for two substances with different microstructures to be completely identical in all of their detectable surface characteristics. A difference in the microstructures of the two substances would have to manifest itself in some detectable way. If this is so, then Putnam has made an illegitimate move at the outset of his discussion when he says: "I shall suppose that XYZ is indistinguishable from water at normal temperatures and

pressures".⁵⁶ He has supposed that a certain scenario is intelligible when in fact it is not. Since the argument for noncognitivism functions by contrasting our linguistic intuitions concerning Twin-Earth scenarios with our linguistic intuitions concerning Moral Twin-Earth scenarios, it presupposes that these scenarios are intelligible. As a result, the argument falls into trouble if it turns out that these scenarios are not intelligible. Unfortunately, discussion of the intelligibility of Putnam's original Twin-Earth scenario is an issue which would send this particular dissertation too far afield. Fortunately, it is not necessary to engage in this kind of discussion. If one is unwilling to accept the intelligibility of the Twin-Earth and Moral Twin-Earth scenarios, essentially the same argument in favor of a noncognitivist characterization of the nature of moral language use can be formulated in terms of a variation of Tyler Burge's discussion of beliefs about arthritis which occur in different linguistic communities. In his discussion, Burge contrasts two worlds which are identical in their physical constitutions and differ only in their linguistic practices. Thus, potential intelligibility problems do not arise in this alternative characterization. Like Putnam, he begins by considering a counterfactual situation which can be described in terms of a pair of scenarios. In the first scenario, a patient we can call Brian has an inflammatory pain

⁵⁶Putnam, 1973, p.309.

in his thigh which causes him to believe that he has arthritis in his thigh. One of the reasons Brian has come to hold this belief is that he lacks expertise in medical matters. He operates with the typical layperson's understanding of medical terminology. He has been previously diagnosed with arthritis in his knee, and since the pain in his thigh feels similar to the pain in his knee, it seems only natural to him to conclude that the pain in his thigh is also due to arthritis. Upon consulting a physician, however, he discovers that this must be a false belief since, by definition, 'arthritis' can only be an ailment which affects the joints. According to Burge, the fact that Brian has an incomplete grasp of the meaning of 'arthritis' plays an important role in determining the content of those of his mental states which pertain to arthritis.

In an ordinary sense, the noun phrases that embed sentential expressions in mentalistic idioms provide the **content** of the mental state or event. We shall call that-clauses and their grammatical variants "**content clauses**". ... When an expression like 'water' [or 'arthritis'] functions in a content clause so that it is not freely exchangeable with all extensionally equivalent expressions, we shall say that it has an **oblique occurrence**. ... [O]bliquely occurring expressions in content clauses are a primary means of identifying a person's intentional mental states or events.⁵⁷

When people have a vague or incomplete understanding of a particular expression, they are often unaware of which other expressions are extensionally equivalent to that expression. We characterize the mental states "whose content involves a

⁵⁷Burge, 1979, pp.,74,76.

notion that the subject incompletely understands"⁵⁸ by using oblique occurrences of that expression to refer to the content of that mental state. Given Brian's incomplete understanding of the nature of arthritis, sentential characterizations of his mental states pertaining to arthritis will contain oblique occurrences of 'arthritis'.

In the second scenario Brian* also has an identical inflammatory pain in his thigh which also causes him to believe that he has arthritis in his thigh. He is no expert in medical matters, either, and operates with the same layperson's understanding of medical terminology as Brian does. He believes that he has arthritis in his knee for the same reason as Brian. He was also previously diagnosed with arthritis in his knee, and since the pain in his thigh feels similar to the pain in his knee, it also seems natural to him to conclude that the pain in his thigh is due to arthritis. When he consults a physician, however, he discovers that this is a true belief since 'arthritis' refers to any painful inflammatory ailment, regardless of its location.

How can Brian and Brian* have identical physical ailments and be in identical brain states, if one of them has a true belief about that ailment and one of them has a false belief about it? The explanation is that although Brian and Brian* are in identical brain states, they each have different beliefs about their identical physical ailments as a result of

⁵⁸Burge, 1979, p.79.

their being in different linguistic communities. Each of them believes that his token of their identical ailment is an instance of the natural kind which is rigidly designated by 'arthritis' in his own community. In each of the two communities, 'arthritis' rigidly designates a different natural kind. So, they have different beliefs. Since they have different beliefs about their identical ailments, it is possible for one of these beliefs to be true while the other one is false. According To Burge:

The upshot of these reflections is that the patient's mental contents differ while his entire physical and non-intentional mental histories, considered in isolation from their social context remain the same. ... The differences seem to stem from differences "outside" the patient considered as an isolated physical organism, causal mechanism or seat of consciousness. The difference in his mental contents is attributable to differences in his social environment.⁵⁹

We can now consider an alternative pair of scenarios. In the first scenario, we have Jeremy B, a man who will not hesitate to transgress the dictates of the categorical imperative whenever this will result in an increase in net global utility. He believes that this behavioral propensity is good. Upon being confronted by the moral experts in his Kantian community, however, he discovers that this is a false belief. They point out to him that a number of his habits are morally unacceptable. He accepts their authoritative declarations on these matters without question since he doesn't have any elaborate criteria for determining the

⁵⁹Burge, 1979, p.79.

extension of 'good' anyway. He just generally tries to act in ways that will make the most people happy. It would be news to him to discover that his behavioral propensity is conducive to maximizing global utility. It would also be news to him that his behavioral propensity transgresses the dictates of the categorical imperative. Being a simple sort, he hasn't really thought about goodness in such terms. Jeremy B is ready to admit that good is a notion that he does not completely understand. Whatever intentional attitudes he has about goodness, they certainly have to be sententially characterized by oblique occurrences of the word 'good'.

In the second scenario, Jeremy B* also has the same behavioral propensity and also believes that this propensity is good. Like Jeremy B, he only has an incomplete understanding of the notion of goodness which can only be properly characterized in oblique terms. He consults the moral experts in his utilitarian community but, unlike Jeremy B, is assured that his belief about this behavioral propensity is true. If 'good' is a natural kind term, then, as in Burge's original characterization, we are led to the conclusion that Jeremy B and Jeremy B* have different beliefs about their identical behavioral propensities as a result of their being in different linguistic communities. Each of them believes that his particular manifestation of their identical behavioral propensity is an instance of the natural kind which is rigidly designated by 'good' in his own community. In each

of the two communities, 'good' rigidly designates a different natural kind. So, they have different beliefs. Since they have different beliefs about their identical behavioral propensities, it is possible for one of these beliefs to be true while the other one is false. This is the only way to account for the fact that Jeremy B*'s belief is true while Jeremy B's belief is false.

Yet, this conclusion does not seem to be warranted in the moral case. The problem with this Burgean scenario, as with the Moral Twin-Earth scenarios, is that it is not possible for Jeremy B to dissent from the moral proclamations of his society. It is not an open question whether or not the characterization of goodness endorsed by the authoritative canons of reasoning accepted by the community is itself good. It has to be good because that is what 'good' means in that community. If we treat a term like 'good' as a straightforward natural kind term, the possibility for legitimate moral disagreement has been curtailed in advance. This makes the project of providing a naturalistic definition of good implausible. This is the same difficulty Moore had with attempts to provide a naturalistic definition of 'good' in his day. It does not matter whether we take the proposed natural definiens to count as the sense or the reference of the definiendum 'good'. The moral theorists of Moore's day attempted to provide a naturalistic characterization of the sense of 'good'. The contemporary moral realists have

attempted to provide a naturalistic characterization of the reference of 'good'. The same problem arises in both cases. The possibility of legitimate moral disagreement is curtailed in advance.

Simple as his moral perspective may be, it certainly seems intelligible for Jeremy B to have a significant moral disagreement with the pronouncements of the moral experts of his society. It certainly seems intelligible for him to question whether or not the authoritative characterization of goodness accepted by the community is indeed good. They cannot, like Brian's doctor in Burge's original scenario, reply that his dissent is nonsensical since it merely betrays his unfamiliarity with the standard usage of the term in question. They will have to provide some argument for the propriety of the standard usage of the term and the impropriety of his deviant usage of the term. In Burge's original scenario, which focuses on the term 'arthritis', the doctor merely has to point out that the term is used in a particular way; he does not have to furnish an additional justificatory story for the appropriateness of that usage. This is the crucial difference between the two cases. The simplest explanation of this difference seems to be that 'good' does not really function as a natural kind term, as the contemporary moral realists have supposed.

The mere fact that Jeremy B's community typically applies the word 'good' to a different complex natural property from

the one that Jeremy B*'s community typically applies the word 'good' to is not enough to show that they are in different intentional states when they claim that their identical behavioral propensity is good. This is because it does not seem accurate to characterize each of them as believing that his particular manifestation of their identical behavioral propensity is an instance of the natural kind which is rigidly designated by 'good' in his own community. It does not seem plausible that the mere fact that Jeremy B is in a different linguistic community than Jeremy B* **makes** his belief that his token of their identical behavioral propensity is good false, and also **makes** Jeremy B*'s belief that his token of their identical behavioral propensity is good true. What we mean by our predications involving a moral term does not depend on the way that the rest of the community happens to make predications involving that moral term in the same way that what we mean by our predications involving a straightforward natural kind term depends on the way that the rest of the community happens to make predications involving that natural kind term. But, if the intentional states that Jeremy B and Jeremy B* are in are properly characterized as cognitive, then they have to be characterized as **different** cognitive states. This gives us a reason to suppose that the intentional states they are in are not cognitive ones. Thus, it gives us a reason for supposing that moral terms are not simply a variety of natural kind terms. In this way, a variation of Burge's

discussion of the natural kind term 'arthritis' can be used to make the same point as the Moral Twin-Earth scenario without encountering potential intelligibility problems.

The second potential problem with using Moral Twin-Earth scenarios as a means of arguing in favor of noncognitivism is that there is an important difference between Putnam's original Twin-Earth scenario and Moral Twin-Earth scenarios. This difference might make it the case that the semantic insights which are illustrated in the former do not hold in the latter. The difference in question is this. In the actual world, we do rigidly designate the substance H^2O by the word 'water' in a univocal fashion. On the other hand, it is not immediately clear that we univocally designate particular classes of natural states of affairs by the word 'just' or 'good' in the actual world. Horgan and Timmons acknowledge this in a footnote where they say:

Indeed, we doubt that there is any single characteristic temperament - any single profile of sentiments - that operates, for Earthlings generally, in matters of morals. But for present purposes one can suppose there is. This supposition fits naturally with the optimistic (though implausible) empirical assumption, which we are already granting for argument's sake, that there is some single set of natural properties that causally regulate the use of moral terms by Earthlings generally.⁶⁰

It seems to me that this difference is not problematic for two reasons. First, if the characterization of moral language use proposed by The Cornell Moral Realists is correct, then

⁶⁰Horgan and Timmons, 1990, fn.12, p.465.

certain classes of natural states of affairs are rigidly designated by terms like 'good' and 'just'. Thus, it will be legitimate to test our linguistic intuitions by comparing plausible characterizations of our putative actual uses of the words 'good' or 'just' with alternative uses of these words on Moral Twin-Earth. Secondly, even if we don't univocally designate particular classes of natural states of affairs with our moral vocabulary, this difference does not affect the ability of the Moral Twin-Earth scenario to illustrate the point that there seems to be a fundamental difference between the way that we use moral terms and the way that we use straightforward natural kind terms. This is all that I am interested in demonstrating at this point. In order to facilitate this, we can suppose, for the sake of argument, that we do univocally designate particular classes of natural states of affairs with our moral vocabulary.

CHAPTER TWO: THE ARGUMENT FROM INTERNALISM

2.1: Introduction

Having presented a preliminary argument in favor of an updated version of noncognitivism in the previous chapter, I will begin this chapter by fleshing out the nature of this updated version of noncognitivism more fully. I will then go on to explain how this updated version of noncognitivism is able to deal with one of the two principal theoretical criticisms of traditional versions of noncognitivism. This is the amoralist argument. There is an important theoretical connection between the amoralist argument and the doctrine of internalism. In the last few years, those interested in metaethical issues have engaged in a fair amount of discussion over the intricacies of internalism. Much of this new discussion is related to the strong anti-internalist position espoused by David Brink. Brink claims that since internalism is incorrect, the only plausible metaethical stance is externalist moral realism. His basis for maintaining that internalism is incorrect centers on the claim that the amoralist is conceivable. Briefly, the argument runs as follows. Noncognitivists are theoretically committed to endorsing the doctrine of internalism. The doctrine of internalism requires that the amoralist is inconceivable. Since the amoralist is conceivable, internalism must be false. Thus, noncognitivism is implausible. These are the grounds

upon which Brink maintains that all forms of noncognitivism are implausible. I will argue that Brink's treatment of this issue is erroneous because it is too narrow in its scope. As a result, his amoralist argument against the plausibility of internalism is not convincing. It does not show that all forms of noncognitivism are incorrect. In particular, it does not show that the version of noncognitivism I am espousing is incorrect. At best it only shows that traditional forms of noncognitivism are theoretically problematic. Thus, the dissertation will take a decidedly defensive tack at this point. Having presented an argument in favor of a particular noncognitivist construal of the nature of moral language use in the previous chapter, in this chapter I shall concentrate on attempting to defend this version of noncognitivism by explaining how it is immune to one of the two standard criticisms of traditional noncognitivism.

2.2: A Three-Level Version Of Noncognitivism

I will start out by emphasizing that I am especially interested in defending the particular version of noncognitivism that I am espousing rather than attempting to provide a generic defense of all versions of noncognitivism. The version I am espousing has three levels. At the highest level, we have noncognitive first-order moral statements (FOMS) such as 'Stealing is bad'. A statement like this is used by members of linguistic communities to express disapproval of that class of actions which their linguistic

community refers to by the word 'stealing'. These statements are most appropriately regarded as noncognitive for the reasons discussed in the previous chapter. Only a noncognitive construal of FOMS allows us to regard a disagreement over the way that a particular FOMS is used as a substantial disagreement rather than a mere instance of speaking at cross-purposes. At the next level, we have cognitive second-order moral statements (SOMS) such as 'Stealing is bad is defensible'. These statements correspond to the intentional states that are presupposed in the sincere utterance of the corresponding FOMS. At the lowest level, we have the ultimate moral perspectives (UMPs) which are embraced by individual users of moral language. An UMP is a pervasive, largely coherent set of noncognitive propensities to react favorably toward certain types of persons, actions, and states of affairs and unfavorably toward others. As an individual goes through life, this set of noncognitive propensities often changes. As a result, so does the precise character of the UMP she embraces. Since societal influences play a significant role in determining the nature of the UMPs that individuals come to internalize, it should come as no surprise that members of the same community often have largely overlapping UMPs. UMPs overlap to the extent that they attribute the same defensibility status to particular FOMS. This overlapping factor allows for the possibility of a significant amount of shared moral ground between individuals embracing different

UMPs. Shared moral ground obtains whenever embracers of different UMPs agree on particular moral issues. For example, if Myron, who embraces UMP_1 , and Selma, who embraces UMP_2 , both agreed that the FOMS 'Pre-marital sex is wrong' is defensible, then they would have shared moral ground on the issue of pre-marital sex. If their whole society agreed with them, then their whole society would have shared moral ground on that issue. If Myron and Selma agreed on almost every moral issue, they would have a large amount of shared moral ground. Their UMPs would almost completely overlap. In this kind of situation, it would make sense to loosely claim that they embraced the same UMP. If their whole society agreed on almost every moral issue, it would make sense to loosely claim that the whole society embraced the same UMP.

Within an UMP, whether a particular FOMS is defensible or not is a factual issue. Thus, it is intelligible to speak of the corresponding SOMS as being true or false within that UMP. What makes the SOMS true, within that UMP, is the fact that the embracer of that UMP is predisposed to endorse the corresponding FOMS. Since she is prepared to endorse the FOMS, by her lights, that FOMS is defensible. Since this constitutes the truth condition for the SOMS within that UMP, the SOMS is true within that UMP. It is also intelligible to speak of a SOMS being true or false within a particular area of shared moral ground. What makes the SOMS true, within that area of shared moral ground, is the fact that the embracers of the

UMPs that occupy that area of shared moral ground are all predisposed to endorse the corresponding FOMS. Since they are prepared to endorse the FOMS, by their lights, that FOMS is defensible. Since this constitutes the truth condition for the SOMS within that area of shared moral ground, the SOMS is true within that area of shared moral ground. As mentioned earlier, when the area of shared moral ground common to a group of individuals becomes pervasive enough, it makes sense to loosely claim that the group embraces the same UMP.

Using this scheme, we can characterize moral language use in the following manner. A FOMS is a linguistic device used by members of various linguistic communities to express their approval or disapproval of various actions, persons, and states of affairs. In expressing such approval or disapproval, a speaker is often making an implicit behavioral demand on others. When he makes this demand, he intends to be taken seriously by the hearers of his FOMS. He expects others to acknowledge that the behavioral demand is defensible. In other words, he expects them to admit that the corresponding SOMS is true. In order to reject this cognitive claim about the FOMS, it is necessary to reject all UMPs that occupy the area of shared moral ground within which the corresponding SOMS is true. Most importantly, it is necessary to reject the UMP embraced by the speaker. As a result of the potentially differing character of the individual UMPs embraced by various participants in moral discourse, FOMS are 'essentially

contestable' utterances. There are three possibilities to consider here. First, in a common range of cases, the FOMS one utters is immediately recognized by its hearer to be defensible within the UMP she embraces. In such cases there is no moral disagreement. There is moral consensus. The FOMS is immediately accepted by its hearer as constituting a defensible behavioral demand. In a second range of cases, the FOMS one utters is not immediately recognized by its hearer to be defensible within the UMP he embraces even though it is. In such cases, there is a potentially resolvable moral disagreement. By appealing to certain intuitions which reflect the pervasive, largely coherent set of noncognitive propensities to react favorably toward certain types of persons, actions, and states of affairs and unfavorably toward others which is constitutive of the UMP of the hearer, it is possible to demonstrate that a FOMS is defensible within an UMP even though its embracer did not immediately recognize this to be so. In a third range of cases, however, the FOMS one utters is not recognized by its hearer to be defensible within the UMP she embraces because it is not, in fact, defensible within that UMP. In such cases, the moral disagreement involved is irresolvable. Since the UMPs embraced by the two parties are so fundamentally different, with respect to the defensibility status of the FOMS in question, there are no intuitions or facts about the world that either of the two parties can appeal to in order to resolve the

disagreement over the defensibility status of that FOMS. Such disagreements can occur between individuals possessing fundamentally different UMPs or between groups of individuals possessing fundamentally different areas of shared moral ground concerning the defensibility status of the FOMS in question. These disagreements cannot even be resolved in principle because the shared moral ground which is lacking is a necessary condition for arriving at a consensus on the defensibility status of the FOMS. This kind of disagreement cannot be resolved through linguistic innovation because its root is fundamentally non-linguistic. The two parties have fundamentally irreconcilable sets of noncognitive propensities which cause their differing moral reactions.

2.3: Two Distinctions Pertaining To Internalism

Having provided a preliminary sketch of the version of noncognitivism I am espousing, I will now set out David Brink's complex defense of externalism which culminates in an appeal to the amoralist argument. This discussion has become one of the standard accounts on this subject. Several contemporary authors recommend Brink's comprehensive discussion to the reader in their treatment of this issue.¹ His account centers on two principal distinctions which allow us to consider several different varieties of internalism. The first distinction concerns the nature of the effect that moral

¹Goldsworthy, 1992, p.40, ftn.1.
Campbell, 1993, p.315, ftn.11.
Little, 1994a, p.151, ftn.4.

considerations have upon agents. There are two possibilities here. One way that a moral consideration can affect an agent is by causing him to have **some motivation** to act in accordance with it. Another way is by providing him with **a reason** to act in accordance with it. The kind of reason intended here is a good or justificatory reason. These are reasons which would be endorsed by fully rational agents.

The second main distinction concerns the perspective from which the relation between moral considerations and reasons or motives is considered. There are three possibilities here. 'Agent Internalism' considers the matter from an objective perspective. According to agent internalism "it is in virtue of the concept of morality that **moral obligations** motivate or provide reason for the agent to do the right thing." Part of **having** an obligation to do x is having a reason or motive to do x. 'Appraiser Internalism' considers the matter from a subjective perspective. According to appraiser internalism "it is in virtue of the concept of morality that **moral belief or moral judgement** provides the appraiser with motive or reason for action." Part of **believing that you have an obligation** to do x is having a reason or motive to do x. 'Hybrid Internalism' combines the two previous forms. According to hybrid internalism "it is a conceptual truth about morality that the **recognition of a moral obligation** motivates or provides the agent with reason for action."² Part of

²Brink, 1989, pp.40,41.

recognizing (having and believing that you have) an obligation is having a reason or motive to do x.

By combining every possible combination of these two distinctions, Brink's scheme allows for six different types of internalism. Keeping this background in mind, we can focus on the task of characterizing the precise variety of internalism that the version of noncognitivism I am espousing is essentially committed to. We can begin by considering the nature of the effect that moral considerations have upon agents. According to my scheme, the sincere endorsement of a FOMS is conceptually tied both to having a reason to comply with its dictates as well as having some motivation to comply with its dictates. It is conceptually tied to having a reason to comply with its dictates since the endorsement of the FOMS presupposes the truth of the SOMS which claims that the FOMS is defensible. To admit that a behavioral demand is defensible is to admit that we have some reason to comply with it. If we had no reason to comply with it, then, by that very token, it wouldn't be defensible. It is conceptually tied to being motivated to comply with its dictates since the sincere endorsement of a FOMS presupposes the embracing of an UMP which holds that FOMS to be defensible. We have a standing motivation to act in accordance with the set of noncognitive evaluative propensities that form an integral part of our personality. We are naturally motivated to act in ways that we

find intuitively appealing and to avoid acting in ways that we find intuitively repulsive.

Brink claims that the two ways moral obligations can affect agents have often been conflated in the literature. According to my characterization of moral language use, this 'conflation' is not a matter of confusion or theoretical carelessness. Rather, it is a reflection of the complex nature of the intentional state involved in assenting to a moral obligation. A person can only take herself to have a moral obligation to do x if she takes herself to embrace an UMP where it is defensible to hold that she has a moral obligation to do x. In such instances she will both have some motivation to do x, and take herself to have a reason to do x. Thus, the two possibilities are simply different sides of the same coin.

At this point, someone might object that a person could be mistaken about herself when she takes herself to have an obligation to do x. For example, a young woman might have been browbeaten into thinking that she had a moral obligation to abstain from alcohol, when she did not, in fact, embrace an UMP where it is defensible to hold that it is wrong to drink alcohol. Later, she might partake of alcohol and experience no guilt or regret for having done so. At this point, she would realize that she no longer takes herself to have a moral obligation to abstain from alcohol. She did take herself to have this obligation at one time, however. Doesn't this show that it is false that a person can only take herself to have

a moral obligation to do x if she embraces an UMP where it is defensible to hold that she has a moral obligation to do x? It seems to me that it does not for the following reason. When she was temporarily browbeaten into thinking that she had a moral obligation to abstain from alcohol, she was browbeaten into temporarily embracing an UMP which she never came to truly embrace. If she had come to truly embrace that UMP, she would have experienced guilt or regret when she later consumed alcohol. Her temporary acceptance of the moral obligation to abstain from alcohol was contingent on her temporary acceptance of an UMP which was not consonant with her more permanent personality traits. When she later found herself partaking of alcohol without any characteristic feelings of guilt or regret, it became evident that she never really internalized the UMP that she was browbeaten into temporarily embracing. At this point, she no longer accepted that she had a moral obligation to abstain from alcohol because she no longer embraced the UMP she was browbeaten into temporarily embracing. Thus, the moral obligations she took herself to have were always a function of the UMP she presently embraced.

We can now consider whether the version of noncognitivism I espouse is committed to agent, appraiser, or hybrid internalism concerning the relation between having a moral obligation and having a reason or motive to act accordingly. Each possibility will be considered in turn. First, my version of noncognitivism is committed to a limited sort of agent

internalism. If, within a particular area of shared moral ground, it is defensible to claim that x has an obligation to do y, then, within that area of shared moral ground, x has a reason to do y. Outside of that shared moral ground, whether or not x has a reason to do y is indeterminate. A parallel situation holds in the case of motivation. If the UMP that x holds is within that range of UMPs where it is defensible to claim that x has an obligation to do y, x will have some motivation to do y. Second, it is committed to appraiser internalism. If x can sincerely assent to her having an obligation to do y, then she is of the view that it is defensible to claim that x has an obligation to do y. This requires that the UMP she presently embraces is within that range of UMP where it is defensible to hold that x has an obligation to do y. It also requires that she take the corresponding SOMS to be true. Thus, for the reasons mentioned in the previous paragraph, if x sincerely assents to having an obligation to do y, x will have some motivation as well as take herself to have some reason to do y. The third alternative, hybrid internalism, presents an odd situation, since, according to the version of noncognitivism I am espousing, one can only 'recognize' an obligation in the sense of recognizing that the behavioral demand corresponding to that obligation is defensible within a certain range of UMPs. For x to 'recognize' that he has a moral obligation to do y is simply for him to manifest the fact that the UMP he embraces

is within that range of UMPs where the behavioral demand corresponding to the obligation is defensible. In such a case, he will have some motivation to do *y* and see himself as having some reason to do *y*, again, for the reasons mentioned in the previous paragraph. Whether she does indeed **have** a reason to do *y*, is not a matter which obtains independently of particular UMPs.

2.4: Brink's Three-Step Characterization Of Internalism

Having made clear the manner in which the version of noncognitivism I am espousing corresponds to Brink's initial two distinctions, we can go on to consider Brink's positive characterization of internalism. He describes internalism as the endorsement of the following three claims:

- (1) Moral considerations **necessarily** motivate or provide reason for action. ...
- (2) Since it is the **concept** of morality that shows that moral considerations necessarily motivate or provide reasons for action, this claim about the motivational power or rationality of morality must be a priori. ...
- (3) Since it is the **concept of morality** that determines this fact, the rationality or motivational power of moral considerations cannot depend on substantive considerations such as what the content of morality turns out to be, facts about agents or the content of the correct theory of rationality. ...

Externalism is the denial of internalism; externalism claims that the motivational force and rationality of moral considerations depends on factors external to the moral considerations themselves.³

As readers familiar with Brink's discussion of these issues will immediately recognize, the most important theoretical upshot of the preceding discussion, by far, is the clear

³Brink, 1989, p.42.

admission that my characterization of noncognitivism is committed to the claim that moral considerations **necessarily** motivate or provide reasons for action since there is a **conceptual connection** between the making of a sincere moral attribution and being motivated or taking oneself to have a reason to act in accordance with the dictates of that moral attribution. Having made this bold admission, I will now sketch the details of the standard theoretical debate which eventually comes to center around the putative implications of this conceptual connection and the necessary connection that accompanies it.

2.5: Internalism, Noncognitivism, And The Amoralist

On the face of it, realism vs. noncognitivism and internalism vs. externalism seem to be entirely distinct debates. One is concerned with issues in the philosophy of language while the other is concerned with issues in action theory. The position one takes on the first issue does not seem to be related to the position one might take on the second issue. Many people have not seen things in this way, however. They believe that there is an important connection between these two debates. As David Solomon notes:

For many philosophers, the central issue that determines the plausibility of moral realism is the question of whether internalism is true. They are driven to this view because they hold that the most powerful argument against realism presupposes the truth of internalism.⁴

⁴Solomon, 1987, p.382.

If one adopts an internalist position, according to this school, moral realism must be abandoned. Inspired by John Mackie, this argument against moral realism can be characterized as follows:

- (1) Internalism is true.
- (2) So, if moral realism is true, it must be internalist moral realism.
- (3) But, if internalist moral realism is true, then 'queer' moral properties which have mysterious intrinsic motivating powers must exist.
- (4) Such properties do not exist.
- (5) So internalist moral realism isn't true.
- (6) So, moral realism isn't true.

There are two possible responses moral realists can make against this kind of argument. One is to concentrate on (3) and (4) by arguing for a viable form of internalist moral realism. People like John McDowell⁵, David McNaughton⁶, and David Wiggins⁷ attempt to describe our moral experience in such a way that the existence of such properties becomes plausible. If their attempt is convincing, then (4) becomes false and the argument is rendered harmless.

The second and more common response to this argument by moral realists is to deny (1). If (1) can be shown to be

⁵McDowell, 1985, pp. 166-180.

⁶McNaughton, 1988.

⁷Wiggins, 1987, pp. 127-165.

false, then the argument is also rendered harmless. This second strategy typically embodies an offensive element as well. The realists offer the following additional argument:

(1) Noncognitivism is necessarily committed to an internalist position.

(2) Internalism is false.

(3) So, noncognitivism is false.

Brink describes the rationale for (1) as follows:

[T]raditional noncognitivists ... must be (appraiser) internalists (about motives). Whatever the differences between emotivism and prescriptivism, both claim that it is an essential part of the meaning of moral judgements to express the appraiser's attitudes or commitments. On both forms of noncognitivism, therefore, it is part of the meaning of moral judgements, and thus a conceptual truth, that the appraiser holds a pro-attitude to things judged moral and a negative attitude to things judged immoral. This builds motivation into the concept of a moral judgment or belief and so commits noncognitivism to internalism about motives.⁸

As a result of this theoretical background, the question of whether or not internalism is true has come to occupy an important place in the debate between moral realists and noncognitivists. If internalism is true, moral realism must be abandoned. If internalism is false, however, noncognitivism must be abandoned.

The crucial realist argument against the plausibility of internalism consists in the presentation of a very simple counterexample: the amoralist. "The amoralist is someone who recognizes the existence of moral considerations and remains

⁸Brink, 1989, pp.83,84.

unmoved."⁹ If the amoralist is possible, then internalism is false. Thus, noncognitivism must be false. On the other hand, if the amoralist counterexample can be dealt with, the principal realist argument against internalism will have been defused. Thus, moral realism will fall prey to the Mackie-style argument. In this way, fairly large issues in moral theory have come to hinge upon what might initially appear to be a rather trivial point: whether or not the amoralist is conceivable.

2.6: The 'Oh Yeah?' Debate

As in any other philosophical dispute, there are two possible responses that the noncognitivist can make when he is presented with the claim that the amoralist is conceivable. These are 'Oh Yeah?' and 'So What?'¹⁰. The traditional response by noncognitivists has been of the 'Oh yeah?' variety. That is to say, traditional noncognitivists have accepted the claim that the conceivability of the amoralist would pose a significant difficulty for the plausibility of noncognitivism and, as a result, have centered their efforts on attempting to establish that a genuine amoralist is not really conceivable. The most common response to the amoralist counterexample by such noncognitivists is to reply that the

⁹Brink, 1989, p.46.

¹⁰This terminology comes from Nicholas Sturgeon. In a footnote, he tells us that he got it "From a tradition among Cornell graduate students. I learned of it from Brian Penrose, whom I discovered analysing my lectures in these terms for undergraduates." Sturgeon, 1986, p.136.

putative amoralist is not making a genuine moral judgement when she 'recognizes' the existence of a moral consideration. Rather, she is simply reporting that the consideration in question is taken to be a moral consideration by the community which she finds herself a part of. As a detached spectator, she can accurately report on which considerations are morally endorsed by the community without morally endorsing such considerations herself. She is using the word 'moral' in an 'inverted commas' sense here. In the words of R. M. Hare, "we are, in this use, not making value judgements ourselves, but alluding to the value judgements of other people."¹¹ The amoralist can accept that she has a moral obligation in this detached sense, but this is not the same as making the genuine moral judgement that she has a moral obligation. Since she is not making a genuine moral judgement, this counterexample does not constitute a refutation of internalism. Thus, traditional noncognitivists have attempted to deal with the amoralist argument by refusing to accept the terms in which it is presented.

The typical response to this tactic by the moral realist is that the noncognitivist is simply begging the question. According to Brink:

It is simply unclear why we should assume that the person who professes indifference to what she insists are moral requirements is confusedly using moral language in inverted commas... We can imagine someone who regards what we take to be moral demands as moral demands - and

¹¹Hare, 1952, p.124.

not simply as conventional moral demands and yet remains unmoved.¹²

It is open to the noncognitivist to respond in kind, however.

As David Solomon points out:

The notion of an amoralist is theory-laden. ... The conceivability of the amoralist cannot be used as a fixed point prior to ethical theory that can be used to justify moves within ethical theory. It is too heavily laden with theory itself to play such a role. That is why the device of pointing to the conceivability of the amoralist has about it a question-begging air. From the standpoint of ethical theory the conceivability of the amoralist is the problem; it cannot be part of the solution.¹³

According to Jeffrey Goldsworthy:

The question is which description of the phenomenon is correct, the externalists' or the internalists'. This question cannot be settled just by considering the phenomenon itself. Brink's proposed thought experiment does not advance the debate. The result is a standoff, externalists asserting that their description of the phenomenon is correct, and the internalists that theirs is. In purporting to 'imagine' genuine amoralism Brink is relying on the very conceptual intuitions whose veracity the internalists deny.¹⁴

Is there any way of resolving the 'Oh Yeah?' standoff?

Perhaps.

The noncognitivist can begin by taking note of the fact that the amoralist is only a fictional character; a figment of the imagination. We are not speaking about actual people here. To be sure, some externalists point to sociopaths as actual examples of amoralists, but even then, they never refer to any concrete, actual persons. They don't need to. According to the

¹²Brink, 1989, pp.47,48.

¹³Solomon, 1987, p.385.

¹⁴Goldsworthy, 1992, p.45.

externalist, the mere fact that we can **conceive** of such a character is enough. Whether or not any actual amoralists exist is beside the point. At this point, the noncognitivist can emphasize the fact that we are all easily able to conceive of fictional characters which cannot possibly exist. The imagination is a powerful thing. We can conceive of an inventor who goes back in time and accidentally kills his own infant father. Yet, we will not admit that such a character is possible. We can conceive of a fictional young student who refutes Gödel on his incompleteness theorem. Yet, experts will refuse to acknowledge that such an event is actually possible. How is the realist to show that his musings about the amoralist are not just another fictional construction of the imagination which is not actually possible? If this is all the amoralist amounts to, he does not present a problem for the internalist. In order to provide a convincing refutation of internalism, the externalist must present a counterexample which is a clear possibility and not just a fanciful chimera of the imagination.

At this point, the externalist might respond that the young student is by no means a clear-cut example of an impossible character. All we can imagine are the surface features of the character. No one could imagine in detail how the character could actually refute Gödel. Here, the noncognitivist will eagerly rejoin that the amoralist is by no means a clear-cut example either. The realist claims that the

amoralist is in a certain complex intentional state which the internalist claims it is impossible for anyone to be in. A crucial problem with this debate is that claims about whether or not people are in particular intentional states do not wear their truth conditions on their sleeves. As a result, we can only make inferences about the character of the intentional states instantiated in various agents. An agent's behavioral manifestations are generally the only clear-cut evidence we can turn to in order to determine whether or not she is in a particular intentional state. Unfortunately, there are no behavioral manifestations available that will conclusively demonstrate the possibility of the peculiar intentional state that is supposedly instantiated in the amoralist. Such behavioral manifestations are only possible given that the possibility of the amoralist is already presupposed. It is not legitimate to **assume** that such behavioral manifestations are possible in order to demonstrate the possibility of the amoralist. Thus, the amoralist counterexample presented by the realist can be legitimately refused by the noncognitivist in a way that a clear-cut counterexample cannot.

As an example of a clear-cut counterexample, we can suppose that, in the heat of an argument about politics and family values, Dan claims: It is inconceivable that a Republican would listen to the Rolling Stones. In response, Bill points out that his neighbor is a registered member of the Republican party who possesses a large collection of

Rolling Stones albums whose sounds can often be heard emanating from his basement window. **This** constitutes a clear-cut counterexample. The two properties are conceptually independent and there are uncontroversial public criteria which establish both that the neighbor is a Republican and that he listens to the Rolling Stones. In this situation, it would not do for Dan to claim that the neighbor wasn't **really** a Republican or didn't **really** listen to the Rolling Stones. He clearly satisfies the public criteria for being the type of person in question. It is an illegitimate move to make an ad hoc change in the criteria after the counterexample has been presented. This is not what is happening in the amoralist counterexample, however. Here, it is by no means clear that the candidate has satisfied the criteria in question because these criteria are not public, nor are they conceptually independent. As Solomon notes, all we have been supplied with is a description of "a particular package of utterances and motivations found together in an agent."¹⁵ This description can only be interpreted as an example of an amoralist by assuming what the counterexample is intended to show; namely, that internalism is false. Therefore, the internalist is not at fault when she rejects the terms in which the 'amoralist' counterexample is presented. The realist has not demonstrated that the amoralist really is the sort of person he claims that he is. On the other hand, the internalist has not cogently

¹⁵Solomon, 1987, p.383.

demonstrated that the amoralist is not the sort of person the realist claims that he is. As far as I can see, this is where the discussion must end. The 'Oh Yeah?' debate over the genuine conceivability of the amoralist is destined to remain in a permanent deadlock. In defending its own perspective on the issue, each of the two sides is only able to appeal to intuitions which the other side clearly doesn't share. As a result, both sides are only preaching to the converted.

2.7: A Crude View Of Conceptual Truth

At this point, we need to consider in more detail the reasons why a noncognitivist might feel the need to become involved in the 'Oh Yeah?' debate in the first place. If it turns out that the plausibility of noncognitivism does not turn on the outcome of that debate anyway, then the whole issue of whether or not the amoralist is genuinely conceivable becomes rather moot. I believe that this is the case, at least for my characterization of noncognitivism. Thus, I am quite content to leave the 'Oh Yeah?' debate at a standstill. In what follows, I will argue that the mere conceivability of the amoralist does not falsify internalism. At least, it does not falsify the version of internalism that my characterization of noncognitivism is essentially committed to. So, it does not show that my characterization of noncognitivism is implausible. As a result, the appropriate response to the conceivability of the amoralist is 'So What?'

We can begin the case for the appropriateness of the 'So What?' response by noting that, interestingly enough, Brink's predictable tendency to multiply distinctions at every turn of every issue curiously lapses when he considers the notions of **conceptual truth** and **necessity** as they apply to the internalism issue. He is quite content to suppose that these ideas are straightforward enough to require no further analysis. His treatment of the notion of conceptual truth is particularly troublesome since his amoralist argument against the plausibility of internalism hinges on the claim that "internalism is the view that there is an internal or conceptual connection between moral considerations and action or the sources of action".¹⁶ The troublesome aspect of Brink's approach is this. Unless he can show that all forms of noncognitivism are committed to a version of internalism which embodies **his kind of conceptual connection** between moral considerations and reasons or motives to act, the amoralist argument fails as a refutation of all forms of noncognitivism. At best, it merely serves to cast doubt on those types of noncognitivism that are committed to a form of internalism which embodies **this** kind of conceptual connection between moral considerations and reasons or motives to act. It cannot function as a basis for rejecting internalism tout court. Thus, it cannot function as a basis for rejecting noncognitivism tout court.

¹⁶Brink, 1989, p.38.

Michael DePaul makes the following comments on Brink's use of the amoralist in this context.

If Brink's aim is to argue against internalism in general, on the grounds that internalists cannot take amoralism seriously, then his argument fails because it is grounded in a crude view of conceptual truths. If all conceptual connections were transparent to anyone possessing the relevant concepts, then perhaps the argument would succeed. It would, however, be quite unnecessary, since if conceptual connections were transparent everyone having the relevant concepts could immediately tell whether the conceptual connection involved in internalism obtains. But conceptual connections are not all transparent. It is often extremely difficult to distinguish what is conceivable from what merely seems so or to determine whether a position is truly coherent.¹⁷

It is significant to notice that the crude version of conceptual truth Brink employs here is the same one that played such a crucial role in Moore's original formulation of the Open Question Argument. In fact, citing the conceivability of the amoralist as a counterexample to the plausibility of internalism is itself a rendition of Moore's original OQA strategy turned against the noncognitivist. It works in the following manner.

For simplicity's sake, we can characterize noncognitivism as the view that the purpose of moral language use is to express approval and disapproval of various types of persons, actions, and states of affairs. If this view is correct, then at the time that someone is sincerely expressing an attitude of approval or disapproval toward some person, action, or state of affairs, she is in a particular motivation-laden,

¹⁷DePaul, 1993, p.532.

noncognitive intentional state. Since the existence of this kind of a motivation-laden, noncognitive intentional state is a necessary condition for a sincere moral utterance, it **cannot** be an open question whether or not any particular instance of a sincere moral utterance will be accompanied by the appropriate motivation-laden, noncognitive intentional state. The existence of the sincere moral utterance **necessitates** the existence of the appropriate motivation-laden noncognitive intentional state. But an externalist like Brink will say that it **is** an open question since it is at least **conceivable** that a sincere moral utterance will not be accompanied by the appropriate motivation-laden, noncognitive intentional state. This is precisely what happens in the case of the amoralist's moral utterances. Any sincere moral utterance that the amoralist makes will not be accompanied by the appropriate motivation-laden, noncognitive intentional state. Thus, the mere conceivability of the amoralist eliminates the possibility of characterizing the content of sincere moral utterances in terms of motivation-laden, noncognitive intentional states. Recall that a key presupposition behind the amoralist argument is the claim that:

Since it is the **concept** of morality that shows that moral considerations necessarily motivate or provide reasons for action, this claim about the motivational power or rationality of morality must be a priori.¹⁸

¹⁸Brink, 1989, p.42.

It is important to notice that not all noncognitivists are obliged to accept this claim. Brink's argument fails precisely because he mistakenly assumes that all versions of noncognitivism are committed to this claim. This mistaken assumption accrues from his employing a crude version of conceptual truth throughout his discussion of internalism. Brink falsely assumes that noncognitivists must treat all conceptual truths pertaining to morality as transparent. As a result, he falsely assumes that noncognitivists must treat all conceptual truths pertaining to morality as knowable a priori. In order to illustrate the problem with the preceding presupposition, we can begin by noticing that Brink cannot intend to apply the preceding claim about the concept of morality to the layperson's concept of morality. It isn't plausible to attempt to apply it to **that** concept. Since the average person doesn't think about moral matters in such abstract terms, his concept of morality need not entail anything about motivation or reasons for action. Rather, Brink must intend to apply this statement to the abstract, theoretical conception of morality proposed by noncognitivists. This is the concept of morality he is attempting to refute. **This** concept of morality is couched in terms of what the noncognitivist takes to be the essential properties of a moral consideration. For the noncognitivist, moral language use is conceived of as a means of expressing approval and disapproval of various types of persons, actions,

and states of affairs. This noncognitivist technical concept of morality entails a necessary connection between moral considerations and motivation or reasons for action because it takes the latter to be an essential feature of the former.

The key problem with Brink's analysis is that he is arguing against the theoretical plausibility of a technical concept on the grounds that the semantic implications of this technical concept don't square with the semantic intuitions that are associated with the layperson's related non-technical concept. As a result, the cogency of his argument rests on an equivocation involving two very distinct, yet related, concepts. On the one hand, we have the noncognitivist's technical concept of morality. This concept is couched in terms of what the noncognitivist takes to be the essential nature of moral discourse. On the other hand, we have the layperson's non-technical concept of morality. This concept simply delineates a practical method for easily identifying instances of moral discourse. As such, it is not couched in terms of any putative essential properties of morality. It is an opaque concept. We should expect that anyone who employs **this** concept of morality will take it to be an open question whether or not a particular sincere moral attribution will be correlated with the motivation-laden, noncognitive intentional state that the noncognitivist's technical concept of morality takes to be an essential property of any sincere moral attribution. Thus, anyone employing **this** concept of morality

will find the amoralist conceivable. But this observation about the layperson's concept of morality does nothing to show that the noncognitivist has not provided an accurate characterization of the nature of a sincere moral attribution when she claims that it is the expression of a particular motivation-laden, noncognitive intentional state.

A general problem with this line of argument can be illustrated with an analogous example in the scientific realm, using the concept of salt. As in the previous case, we will not be considering the layperson's non-technical concept of salt. Rather, we will be considering the technical, revisionist concept of salt employed by the modern scientific community. This definition is couched in terms of the essential properties of salt. 'Salt' is defined as follows:

A crystalline compound NaCl that is the chloride of sodium¹⁹.

Consider the following claim about the technical, scientific concept of salt which parallels Brink's claim about the noncognitivist's technical concept of morality:

Since it is the **concept** of salt that shows that salt is necessarily composed of NaCl, this claim about the NaCl composition of salt must be a priori.

What conclusions can we draw about the suitability of this scientific concept by considering the way that it compares to the layperson's concept? None. The layperson's non-technical

¹⁹Webster's Seventh New Collegiate Dictionary.

concept of salt is of a white grainy substance with a certain characteristic taste which can enhance the taste of certain foods when used in moderation. This concept simply delineates a practical method for easily identifying the presence of salt. As such, it is not couched in terms of any putative essential properties of salt. It, too, is an opaque concept. Thus, we expect that anyone who employs the layperson's concept of salt will take it to be an open question whether or not a particular sample of salt is comprised of the chemical compound NaCl. But this observation does nothing to impugn the veracity of the technical, scientific concept of salt which takes NaCl to be an essential property of salt. In the cases of both 'salt' and 'morality', the semantic implications of a non-technical concept cannot be used as a basis for criticizing the theoretical adequacy of a technical concept which as a matter of a posteriori contingency turns out to be co-extensive with the non-technical concept. The amoralist argument doesn't work for the same reason that Moore's original formulation of the OQA doesn't work. It relies on the presupposition that our concepts are transparent. As a result, it relies on the presupposition that all truths about the essential natures of the referents of our concepts can be known a priori. It is important to notice that the noncognitivist is not essentially committed to either of these claims.

2.8: Conceptual Truth And Moore's Original OQA

In the light of this criticism, one might wonder why Brink and others are so willing to foist such an obsolete, implausible doctrine of the nature of theoretical conceptual truth on the noncognitivist, and then proceed to argue against her as if she is essentially committed to it. The reason is that traditional noncognitivists have relied on Moore's original OQA, and the crude version of conceptual truth that goes with it, to furnish the grounds for rejecting moral realism. As a result, they are committed to taking this dubious conceptual baggage along with them. They are open to arguments against internalism, like Brink's, which employ the crude version of conceptual truth that they have already accepted as part of their case for rejecting moral realism. Since these versions of noncognitivism are already committed to a particular characterization of the nature of conceptual truth, they **are** committed to a form of internalism which embodies the kind of conceptual connection between moral considerations and reasons or motives to act that Brink presupposes in adducing the conceivability of the amoralist as a counterexample to the plausibility of internalism. We might say that the amoralist argument constitutes an 'internal' criticism of the kind of internalism that traditional noncognitivists are committed to. Generally speaking, it's a bad argument, but it still works against the traditional noncognitivist. The contemporary moral realist might not accept the characterization of conceptual truth employed in

the amoralist argument, but the old-style noncognitivist has to accept it since she presupposes it in her rejection of moral realism. So the amoralist argument still constitutes a valid objection to her position. At this point, it is important to notice that the version of noncognitivism I am defending is not open to this internal criticism of internalism because it is not committed to the obsolete doctrine of conceptual truth that plays such an important role in Moore's original formulation of the OQA. My rejection of moral realism is based on the cogency of the Moral Twin-Earth argument. It is crucial to notice that the Moral Twin-Earth argument does **not** presuppose the crude version of conceptual truth we find in the original OQA.

2.9: Two Kinds Of Necessary Connection

Having made this point clear, we can now consider two distinct kinds of necessary connection that are muddled together by the crude version of conceptual truth Brink employs in his discussion of internalism. This distinction will play an important role in explaining why the mere conceivability of the amoralist need not present a problem for noncognitivism. I will begin by describing the nature of the necessary connection between moral considerations and motivation or reasons for action that old-style noncognitivists are committed to. As a result of basing their theories on Moore's original version of the OQA, traditional noncognitivists are committed to claiming that it is **analytic**

that moral considerations necessarily motivate or provide reason for action. As a result, they are committed to defending a de dicto necessary connection between moral considerations and motivation or reasons for action. This position can be symbolized as follows:

$$\Box(\forall x)[(x \text{ is a moral consideration}) \rightarrow (x \text{ motivates or provides reason for action})].^{20}$$

If this position is correct, we can know a priori that moral considerations necessarily motivate or provide reason for action simply by analysing the concept of a moral consideration. According to this position, motivation or reason for action are constitutive aspects of the concept of morality. As a result, just as we can know a priori that the concept of a round square is unintelligible, we can also know a priori that the concept of moral considerations that provide no motivation or reason for action are unintelligible. Thus, we can know a priori that the amoralist is unintelligible. The fact that the amoralist *seems* to be intelligible presents a problem for this position. Since the notion of the amoralist is intelligible in a way that the notion of a round square is not, the de dicto claim that moral considerations necessarily motivate is not plausible. Since the traditional noncognitivist is committed to defending this de dicto claim, the intelligibility of the amoralist poses a substantial problem for her position.

²⁰I adopt this symbolization scheme from Brower, 1993, p.236.

The important thing to notice is that the contemporary noncognitivist, who bases his rejection of moral realism on the cogency of the Moral Twin-Earth argument, is not committed to the claim that it is **analytic** that moral considerations necessarily motivate or provide reasons for action. For him, noncognitivism can also be construed as the claim that it is a constitutive aspect of the discursive activity we call **moral** discourse that moral considerations necessarily motivate or provide reasons for action. This type of noncognitivist is only committed to defending a de re necessary connection between moral considerations and motivation or reasons for action. This position can be symbolized as follows:

$$(\forall x)[(x \text{ is a moral consideration}) \rightarrow \Box(x \text{ motivates or provides reason for action})].$$

Even if we assume that this position is correct, it does not follow that it can be known a priori that moral considerations necessarily motivate or provide reasons for action. We cannot know this simply by analysing the concept of a moral consideration. We can only discover a posteriori that this necessary connection obtains as a result of a posteriori discoveries about the nature of the discursive activity we have come to rigidly designate as **moral** discourse. Once this a posteriori determination has been made, the essential nature of the discourse we call 'moral' discourse becomes the basis for maintaining that a necessary connection between moral considerations and motivation or reasons for action obtains; not the meaning of the word 'moral'. Thus, it is quite

possible for someone to know the meaning of the word 'moral', yet be unaware of certain aspects of the essential nature of moral discourse. As a result, it is possible for moral considerations to be de re necessarily connected to motivation or reasons for action while the typical moral language user conceives of this not being the case. Thus, the de re claim that moral considerations are necessarily connected to motivations or reasons for action is compatible with the conceivability of the amoralist. So the mere conceivability of the amoralist does not automatically refute this characterization of noncognitivism.

2.10: Moral Twin-Earth Reconsidered

The additional problem faced by this characterization of internalism, of course, is that the noncognitivist takes on the onus of providing grounds for the claim that she has provided an accurate a posteriori characterization of the essential nature of the discursive activity we rigidly designate as **moral** discourse. Interestingly enough, the amoralist will play an important role in the establishment of this a posteriori case. I will now turn to the task of furnishing such a posteriori grounds. In order to accomplish this, it will be helpful to start by emphasizing an important semantic presupposition that underlies the Moral Twin-Earth argument against moral realism. Following Putnam's original Twin-Earth argument, the Moral Twin-Earth argument assumes that the meanings of words can only be adequately cashed out

in terms of the way that they are used by linguistic communities. This perspective is a reflection of the tendency towards non-solipsistic semantics which has its roots in Wittgenstein's presentation of the private language argument. According to this kind of approach to meaning, conceptual connections are cashed out in terms of the corrective tendencies of the linguistic community. Members of linguistic communities who employ deviant uses of terms will be subject to linguistic censure by the rest of the linguistic community. The **real** meaning of any term is determined by the general tendency of the linguistic community to **use** that term in a particular manner. Within such a framework, whether or not two terms are conceptually connected is a matter which can only be determined a posteriori. We need to determine how a linguistic community uses the terms in question before we can make a ruling on whether or not they are conceptually connected. Thus, the possibility of establishing conceptual connections between two terms on an entirely a priori basis is, in an important sense, no longer possible once a solipsistic approach to semantics is rejected for the simple reason that the community must be consulted in order to verify the way that each of the terms are actually used. This holds for conceptual connections involving moral or non-moral terms. To adopt a more traditional Wittgensteinian parlance, the employment of moral terminology, like any other terminology, takes place within a particular language game. To use moral

language is to act as a participant in this language game. The rules for the appropriate usage of moral terminology are constitutive of the **practice** of moral language use. Moral language use is a particular **type** of language game. What is going to make moral language use distinctly 'moral' will be the specific character of the set of standards which the community understands to govern the appropriate use of moral language.

What the Moral Twin-Earth argument shows is that there is a peculiar difficulty involved with attempting to characterize the meaning of moral expressions in terms of a causal theory of reference. This difficulty is intimately linked with the doctrine of internalism. The reason we take Earthlings and Moral Twin-Earthlings to be having a substantial disagreement over the nature of justice is that we are implicitly presupposing that a particular version of internalism is true. We take it as given that, in general, people have a substantial degree of concern over which types of complex natural states of affairs **moral** terms like 'justice' are applied to. It is important to notice that this is not an a priori stipulation about the meaning of the word 'moral'. Rather, it is an a posteriori observation about the essential nature of that type of discursive activity which we **recognize** to be of the moral variety. The type of discourse that we have come to rigidly designate as **moral** discourse occurs when members of communities apply moral terminology to persons,

actions, and states of affairs as a means of expressing their approval or disapproval of the persons, actions, or states of affairs in question. Since people naturally have some concern over whether or not the things they approve or disapprove of obtain, it stands to reason that moral discourse is **intrinsically connected** to people's concerns. Thus, it stands to reason that there is an intrinsic connection between making a moral attribution and taking oneself to have a reason to act in a way that conforms to the dictates of that moral attribution.

This might not have been the case, however. It could have turned out that people were only **incidentally** concerned with the upshot of their moral attributions because the essential purpose of making moral attributions was not to express approval or disapproval. It could have turned out that communities used moral language simply as a means of describing and classifying various types of persons, actions, and states of affairs in accordance with traditions they had long since ceased to care about. If this were the case, then the type of discursive activity that would have come to have been rigidly designated as **moral** discursive activity would have been of a purely classificatory nature. This type of discursive activity does not require any intrinsic connection between the making of moral attributions and people's concerns or reasons for action. It would be **moral** discourse purely in

virtue of the particular subject matter that it incorporated into its classificatory scheme.

The point to be emphasized here is that determining the correct characterization of 'moral', as a word which denotes a specific type of discursive activity, is an issue that can only be resolved on a posteriori grounds. We can't tell in advance what **moral** discourse is until we find out what typical language users take people to be doing when they are engaging in **moral** discourse. It makes no sense to stipulate in advance that 'moral' is going to designate a particular kind of discursive activity, be it descriptive or prescriptive. Brink, in adducing the amoralist as a counterexample to the plausibility of internalism, and, thus, noncognitivism, is under the mistaken impression that the noncognitivist characterization of the meaning of 'moral' is necessarily based on precisely this kind of a priori stipulation. As previously noted, this is undoubtedly linked to the historical fact that traditional noncognitivists were motivated to advocate their particular characterizations of the meaning of 'moral' as a result of their being convinced by Moore's original formulation of the OQA. They were basing the meaning of 'moral' on a priori considerations. But, a noncognitivist is only committed to a purely a priori characterization of the nature of **moral** language use if his basis for espousing noncognitivism over moral realism rests on an acceptance of something close to Moore's original formulation of the OQA.

Since my rejection of moral realism is based on the cogency of the Moral Twin-Earth argument, the version of noncognitivism I am espousing is not committed to an a priori characterization of the nature of moral language use.

It is important to emphasize that the employment of the Moral Twin-Earth argument itself reflects a commitment to an a posteriori semantics à la Wittgenstein. According to this perspective, the sociological fact that people do find the Moral Twin-Earth argument convincing creates an important a posteriori difficulty for externalist moral realist semantics. If externalist moral realism really does provide an adequate characterization of the nature of our moral language use, our linguistic intuitions would not lead us to endorse the claim that the parties are having a legitimate disagreement when they disagree about whether or not a particular natural state of affairs is good or just. Our linguistic intuitions would lead us to endorse the claim that they are merely arguing at cross purposes in the same way that our linguistic intuitions lead us to endorse the claim that the parties in Putnam's original Twin-Earth scenario are merely arguing at cross purposes when they disagree about whether a particular sample of liquid is water or not. It is crucial to notice that the way this argument will affect people's linguistic intuitions is itself not an issue that can be determined a priori. It could have been the case that no one but a few die-hard noncognitivists found the Moral Twin-Earth argument even

remotely plausible. If this **were** the case, then we would have strong a posteriori grounds for maintaining that moral terms are just another variety of natural kind terms. Since it is not the case, however, we have strong a posteriori grounds for maintaining that moral terms are not just another variety of natural kind terms. Whatever the outcome, determining what human beings are up to when they use moral language is an issue which can only be resolved on a posteriori grounds.

2.11: Moral Language Use As Essentially Group-Oriented

In order to facilitate the resolution of the a posteriori question concerning what actual linguistic communities take **moral** language use to consist in, it is important to focus on a prominent feature of actual moral language use. Actual moral discourse is an intrinsically group-oriented activity. Consideration of this feature will have significant implications for the correct a posteriori determination of the nature of moral language use for the following reason. This feature plays an important role in explaining why the externalist moral realist cannot simply formulate an updated version of the amoralist argument. In the spirit of this new approach, the contemporary moral realist might claim that it is a significant a posteriori fact that typical actual moral language users seem to find no difficulty in conceiving of the amoralist. Doesn't this observation constitute a posteriori grounds for claiming that noncognitivism has not provided an accurate characterization of the essential properties of

actual moral language use? The answer is no, because actual moral language use is a group-oriented activity. A major problem with the amoralist argument is that it ignores the communal aspect of moral language use. The focus of this argument is entirely on the nature of the intentional states of **individual** language users in their private employment of moral terminology. The fact that we can conceive of an individual moral language user, the amoralist, who sincerely asserts that stealing is wrong yet is not motivated to act in accordance with this assertion because she does not take herself to have a reason to do so, supposedly reveals an important insight into the essential nature of moral terminology. It supposedly shows that making a sincere moral assertion and taking oneself to have a reason to act in accordance with the dictates of that assertion are two entirely distinct intentional states which are only accidentally conjoined in the minds of those who are already predisposed to view their own sincere moral assertions in a certain way. Thus, there is nothing in the notion of making a moral attribution **itself** which necessitates taking oneself to have a reason to comply with its dictates. We have to go beyond the mere making of a moral attribution and look for an additional evaluative attitude toward the significance of making of such an attribution, in the mind of the attributor, before we can be assured that a reason to comply with its dictates will also be part of her intentional quiver. The

typical prima facie willingness to comply with the dictates of one's sincere moral attributions, which is not a necessary condition for the making of a moral attribution, functions as the accidental glue between the making of a moral attribution and taking oneself to have a reason to comply with it. When this typical additional element is removed, as it is in the case of the amoralist, it becomes clear that there is no essential connection between making a moral attribution and taking oneself to have a reason to comply with its dictates. Since internalism requires an essential connection between making a moral attribution and taking oneself to have a reason to comply with its dictates, internalism must be false. Since noncognitivism requires that internalism is true, noncognitivism must be false.

The correct line of response by the noncognitivist here will be to argue that this prima facie willingness to comply with the dictates of one's sincere moral attributions is hardly an **accidental** feature which just happens to accompany the overwhelming majority of sincere moral attributions. Rather, it is an **essential** feature which plays an important part in making such attributions **moral** attributions. As such, it functions as the **essential** glue between the making of a sincere moral attribution and taking oneself to have a reason to comply with its dictates. The important point to notice, however, is that it is not necessary for the noncognitivist to demonstrate that this attitude is present in every putative

instance of moral language use in order to show that it is an essential feature of moral language use. It is only necessary for the noncognitivist to demonstrate that this attitude must be a typical feature of the discursive activity we call **moral** discourse in order to show that it is an essential feature of moral language use. I will now make use of an interesting point raised by Richard Miller in order to demonstrate that the latter is indeed the case. He provides a key insight into why it is entirely appropriate for someone who espouses the kind of noncognitivism I have been describing to respond to the conceivability of the amoralist with an emphatic 'So What?'. Miller speaks in terms of the moral nihilist, but the character he is considering is exactly the same as the amoralist. For him, the moral nihilist is a person who does not believe that we have a reason to act morally and, as a result, does not take a moral attribution to provide any reason for action. According to Miller:

The moral nihilist is a semantic parasite. His grasp of the content of moral judgments wholly consists of his understanding of the basis on which others who sometimes do attribute moral terms intend to attribute them. Those others will, at least to some extent, be guided by moral attributions in their projects. As one consequence of this semantic parasitism, there could not be a self-contained community in which someone used terms to refer to moral properties yet no one took the applicability of a moral term to be, in any circumstance, a reason to make a choice. In such a community "wrong" might be used to refer to hurtful acts, acts which we would judge wrong, but the ascription would no more be an ascription of a moral property than "Delicious" as used by agronomists

ascribes delicious taste if it happens to be ascribed to a delicious fruit.²¹

If it is indeed the case that there is no intrinsic semantic connection between the making of a moral attribution and taking oneself to have a reason to act in accordance with its dictates, then it should be possible to conceive of an entire linguistic community where no one sees any connection between the making of a moral attribution and taking oneself and others to have a reason to act in accordance with its dictates. It would be an odd community, to be sure, but if externalism is correct, the possibility of such a community is completely intelligible. Since the externalist denies that there is any intrinsic connection between moral considerations and motivation or reasons for action, she is required to deny that this connection obtains at either the micro or the macro level. As a result, the externalist is committed to the claim that it is possible to have an entire community of amoralists engaging in genuine moral discourse.

The problem, however, is that the type of discourse that the amoralist community engages in when it uses moral terms to merely describe various complex natural states of affairs will not strike actual moral language users as the same kind of discursive practice that we are engaged in when we apply moral terms to various complex natural states of affairs. The key point here is that the fact that the vast majority of moral

²¹Miller, 1992, p.164.

language users in the actual world **do** take their moral attributions to imply the existence of reasons for action has some important bearing on the a posteriori determination of which types of discursive activity are going to count as **moral** discourse. It is also important to notice that it is not necessary to convince the die-hard externalist of the veracity of this position. The mere fact that the vast majority of moral language users conceive of the practice of moral language use in this way is sufficient in itself to vindicate this characterization of the nature of moral language use. The linguistic community, in virtue of who they are, cannot be mistaken on this issue. The word 'moral', as a way of characterizing types of discourse, simply **means** what the vast majority of the linguistic community takes it to mean. It applies to those types of discourse that the vast majority of the linguistic community takes it to apply to. The fallback position of 'I'm still not convinced' is simply not open to the externalist here. Such a refusal to accept this a posteriori characterization of the nature of **moral** language use merely betrays the fact that the externalist is out of step with the rest of the linguistic community concerning the meaning of the word 'moral'. Of course, the externalist is free to characterize the meaning of the word 'moral' in her own idiosyncratic way, but she must realize that this is a deviant characterization. As such, the linguistic intuitions that accompany such a deviant characterization are, for that

very reason, not reliable indicators of what the word 'moral' actually means.

2.12: The Deviant Chess Player

In order to illustrate this point, we can consider the possibility of deviant participants involved in other kinds of practices which require shared understandings. For example, we can imagine Paula, a deviant chess player who is well versed in the strategy of the game, but thinks it is a complete bore. She detests chess and everything to do with chess. Unfortunately, from an early age, she has been forced to participate in the weekly meetings of the local chess club by her well-meaning parents who view this activity as a valuable means of cultivating reasoning and strategic skills in the developing mind of their precious young child. For some time now, Paula has reasoned that if she were to just play the game badly enough, for a long enough period of time, her parents will eventually come to believe that their beloved daughter is just not cut out for chess and will finally put an end her tedium. In such a situation, it should be fairly clear that Paula will not take the observation that she can force a checkmate by implementing a particular series of moves to constitute a good reason for implementing that series of moves. She will not be motivated in the slightest to implement that series of moves because doing that would bring about precisely the kind of outcome she does not want. She doesn't want to succeed at chess. That would only perpetuate her

forced participation in an activity she detests. Of course, it might be the case that she would find herself enjoying chess if she only kept on with it. But that possibility does not matter now. She hates it now and, as a result, has no present motivation and presently takes herself as having no reason to implement a series of moves that will force a checkmate whenever she notices that she has the opportunity to do so.

All of this seems intelligible enough. Chances are that people like Paula are not only possible, but actual. The question is, what does the mere conceivability of a person like Paula show about the claim that, within the game of chess, being able to implement a series of moves that will result in a checkmate is **conceptually connected** with being motivated or having a reason to implement that series of moves? Does the intelligibility of a person like Paula show that these things are not conceptually connected? It would seem not. The game of chess consists of a strategic situation where each of two parties attempts to force a checkmate on the other by moving a set of pieces in a manner that conforms to a certain well-defined set of rules. This is what it **means** to play chess. Chess is an activity which presupposes a certain shared understanding between the two participants. Each of them takes the observation that she can force a checkmate by implementing a particular series of moves to constitute a good reason for implementing that series of moves. If someone attempts to force a checkmate by moving his pieces in a manner

that is not conducive to achieving this goal, we say that he is playing chess badly. It is important to notice that it is not appropriate to characterize Paula as playing chess badly. The problem with the manner in which Paula manipulates chess pieces is not that she lacks the strategic skill which is necessary for the successful execution of checkmates. This is the problem encountered by people who play chess badly. They are trying to accomplish a particular goal, but lack the requisite capacity to do this successfully. Paula does not lack this requisite capacity. She is engaged in a distinctly different kind of goal-oriented activity from the one that the bad chess player is engaged in. She is very good at achieving her goal; he is bad at achieving his. Since the bad chess player's goal is to produce checkmates, he has serious misgivings about his performance. Since Paula's goal is to avoid producing checkmates, however, she has no misgivings about her performance. She is 'playing a different game'. Since it is her implicit intention to avoid producing a checkmate at all costs, it is more accurate to characterize her as not playing chess at all. She is simply moving her pieces in conformity with the rules, always ensuring that her opponent prevails.

To demonstrate this, we can suppose that a particular chess club consists of a whole community of people who all play chess the same way that Paula does. For various reasons, no one in that chess club ever wants to actually win a chess

match. It would not be possible for any two members of this community to engage in an actual game of chess as long as they all held this shared attitude. If they were forced to play, the result would always be an endless series of legal moves by both players which never resulted in a checkmate. No one would ever take the observation that he could force a checkmate by implementing a particular series of moves to constitute a good reason for implementing that series of moves. In short, chess would simply not be possible in such a community. The reason for this is that playing with the intention of achieving a checkmate is a **constitutive aspect** of the game of chess. It is an essential feature of the game that we rigidly designate by the word 'chess'. Any practice which does not include this feature is not the same kind of practice as the one we rigidly designate by the word 'chess' regardless of how similar to chess it may appear to the casual observer. Thus, within the game of chess, recognizing that a series of moves will result in a checkmate is **conceptually connected** with being motivated or having a reason to implement that series of moves. Moreover, it is the kind of conceptual connection that the intelligibility of a deviant counterexample like Paula does not serve to falsify.

The necessary connection within the game of chess between being able to implement a series of moves that will result in a checkmate and being motivated or having a reason to implement that series of moves is an a posteriori conceptual

connection which describes an essential feature of an actual communal practice. It might not have been the case that the game we call 'chess' had this as an essential feature. Since it does, however, this feature must be instantiated in any community in order for its members to be engaging in **that** practice. The same holds for individuals. Any participant who does not instantiate that connection is **by that very token** not engaging in the same practice that the community of participants who do instantiate that connection are engaging in. This feature is shown to be essential by considering an alternate community of participants none of which instantiate that connection. If the practice that the alternate community engages in is of a different kind from the one that the original community engages in, then the feature in question has to be an essential aspect of the practice that the original community is engaged in. This provides a principled basis for characterizing deviant participants such as Paula **as deviant**. She lacks a feature which is essential for engaging in the same kind of practice that the others who do not lack that feature are engaging in. For this reason, it is appropriate to characterize her as deviant. For this reason, it is appropriate to claim that she is not really playing chess.

2.13: The Amoralist As Deviant Moral Language User

The preceding considerations can be transposed to the moral case. We can imagine a person who has mastered the rules

and conventions associated with the canons for appropriate moral language use in his community, but never takes himself to have a reason to act in accordance with the dictates of any moral attribution. This is the amoralist. We can even imagine the amoralist employing sophisticated patterns of moral reasoning. It is also conceivable that an amoralist could be a competent source of moral advice. In giving competent moral advice, the amoralist would be making what Joseph Raz has called 'statements from a point of view'. To make a statement from a point of view is to describe what actions are called for or forbidden from within a particular perspective while remaining neutral on the status of the perspective itself.

It is to be found whenever a person advises or informs another on his normative situation in contexts which make it clear that the advice or information is given from a point of view or on the basis of certain assumptions which are not necessarily shared by the speaker.

Imagine an orthodox but relatively ill-informed Jew who asks the advice of his friend who is Catholic but an expert in rabbinical law. 'What should I do?' he asks, clearly meaning what should I do according to *my* religion, not yours. The friend tells him that he should do so and so. The point is that both know that this is not what the friend thinks that he really ought to do. The friend is simply stating how things are from the Jewish Orthodox point of view.²²

Within a particular perspective, whether or not a course of action should be carried out is a factual matter. In terms of my characterization of noncognitivism, whether or not a FOMS endorsing or forbidding a particular course of action is defensible within a particular UMP or area of shared moral

²²Raz, 1979, p.156.

ground is a factual issue. But it is a factual issue which may or may not be obvious depending on the depth of one's understanding of the perspective and its implications. The important point is that it is not necessary to accept a particular perspective in order to be able to recommend a particular course of action relative to that perspective. To endorse an action within a perspective is to claim that there is a reason to perform that action given that the perspective in question is accepted. It is to claim that there is a relationship between accepting that perspective and accepting that a certain class of actions are required or forbidden. Thus, even an amoralist who is capable of sophisticated patterns of moral reasoning is certainly conceivable.

The question here is, what does the conceivability of the amoralist show about the claim that the making of a sincere moral attribution is **conceptually connected** with being motivated or taking oneself to have a reason to act in accordance with the dictates of that moral attribution? In order to answer this question, we will have to consider what **moral discourse** is. Here, we face a difficulty that does not arise in the chess example. We do not have a pre-existing, well-defined set of rules which delineate the essential nature of moral discourse as we do in the case of the game of chess. In order to discover whether or not a particular characterization of the essential nature of moral discourse is an accurate one, we will have to consult the linguistic

intuitions of actual moral language users. According to my characterization, it is a communal activity whereby members of various communities use the medium of language to praise or criticize certain kinds of persons, actions, and states of affairs. By criticizing the behavior of others in this way, a person using moral language is placing behavioral demands upon them.²³ In making such behavioral demands, she presupposes the legitimacy of an UMP within which such behavioral demands are defensible. If this characterization is correct, there is a definite **conceptual connection** between making a sincere moral attribution and being motivated or taking oneself to have a reason to act in accordance with the dictates of that attribution. Indeed, the presence of this feature is an essential property of the discursive activity we have rigidly designated as **moral discourse**.

In order to test the adequacy of this characterization, we can consider our intuitions concerning the nature of the discursive activity engaged in by an entire community of amoralists. In such a community, statements containing moral predicates are likely to occur much less frequently than they do in our community for the simple reason that there isn't really much point in making them. Since the acceptance of a moral attribution could never count as the acceptance of a reason for action, the making of a moral attribution could

²³Indeed, normal moral language users would take the amoralist to be making such demands were he to employ moral language in their presence.

never serve as the basis for criticizing someone's actions. To inform someone that a particular action is bad or wrong would have about as much effect as informing them that the action in question would not be approved of by Ronald Reagan. The typical reply would be 'So?'. In such a community, no one would ever take someone making a moral attribution to be committed to the claim that people have a reason to act or refrain from acting in a particular manner. In that community, an action's being wrong or a bad thing to do would not count as even a *prima facie* reason to refrain from acting in that manner. Moral language use would be a purely descriptive, if not perfectly idle, activity.

The question we need to consider here is whether the 'moral' discourse that the community of amoralists engages in is of the same kind that we engage in. Are they doing pretty much the same thing with their moral talk as we are doing with ours, or are they doing something quite different? It seems to me to be fairly obvious that they are engaged in a type of discursive activity which is crucially different from the moral discourse that we engage in. It also seems fairly obvious to me that the overwhelming majority of moral language users will agree with me on this point. Notice that this is an *a posteriori* claim about the nature of moral language users in the actual world. As such, it could turn out to be false on empirical grounds. As in the Moral Twin-Earth argument, it might have been the case that no one but a few die-hard

noncognitivists found this whole line of reasoning to be even remotely plausible. This is the a posteriori result we would expect if the characterization of moral language use provided by externalist moral realists is accurate. The hard line on externalism that they embrace commits them to admitting that an entire community of amoralists could still engage in genuine moral language use. I think we will find, as we consult the intuitions of various actual moral language users, that this result does not obtain. Thus, we have a posteriori grounds for rejecting the characterization of moral language use proposed by externalist moral realists in favor of a noncognitivist characterization of moral language. In this way, the noncognitivist can satisfy the onus of providing grounds for the claim that she has provided an accurate a posteriori characterization of the essential nature of the discursive activity we rigidly designate as **moral** discourse. Thus, the de re claim that

$$(\forall x)[(x \text{ is a moral consideration}) \rightarrow \Box(x \text{ motivates or provides reason for action})]$$

is vindicated. This claim can then be directly applied to the amoralist argument.

The problem with the amoralist is not so much that he is using moral terms in a variant 'inverted commas' sense, as that he is engaged in a different kind of practice when he employs moral vocabulary from the one that normal moral language users are engaged in when they employ moral vocabulary. It is not so much what his moral words refer to as

the significance he attaches to their application that makes him a deviant moral language user. It is important to notice that this might not have been the case. It might have been the case that the typical moral language user attached the same significance to her application of moral terminology as the amoralist does. If this were the case, then it would no longer be appropriate to characterize the amoralist as a deviant moral language user. Or it might have been the case that the concern which typically accompanies the application of moral terminology is only an **accidental** feature of moral language use. If this were the case, our linguistic intuitions concerning the nature of the discursive activity engaged in by the community of amoralists would be different from what they actually are. It would no longer be appropriate to characterize the amoralist as engaging in a different kind of discursive activity than the normal moral language user. Whether it is appropriate or not is a matter which can only be resolved by consulting people's intuitions concerning what they actually understand themselves to be doing when they engage in moral discourse. It can't be resolved on the basis of a priori musings in the philosopher's study over what is and what is not conceivable.

An important semantic implication accrues from this a posteriori discovery about the nature of the linguistic intuitions possessed by actual moral language users. If we take the view that when the members of the community of

amoralists employ moral terminology they are engaging in a different kind of discursive activity from the one we are engaging in when we employ moral terminology, then we are committed to the view that there is a **conceptual connection** between the making of a sincere moral attribution and being motivated or taking oneself and others to have a reason to act in accordance with the dictates of that moral attribution. The basis of this conceptual connection is a posteriori, rather than a priori. The conceptual connection obtains as a result of our taking being motivated or taking oneself and others to have a reason to act in accordance with the dictates of a moral attribution to be an **essential feature** of that type of intrinsically communal discursive activity that we rigidly designate as **moral** discourse. As in the chess example, then, we have a principled basis for characterizing the amoralist as **deviant** vis a vis the rest of the community of moral language users. The discursive activity that the amoralist is engaging in lacks an essential feature of moral discourse. Thus, the mere conceivability of the amoralist does not falsify the a posteriori claim that there is a conceptual connection between the making of a sincere moral attribution and being motivated or taking oneself to have a reason to act in accordance with the dictates of that moral attribution. Thus, the conceivability of the amoralist does not refute the doctrine of internalism. Thus, the conceivability of the amoralist does not falsify noncognitivism. Thus, for the noncognitivist, the

appropriate response to the conceivability of the amoralist is 'So What?'. At least, this is the appropriate response for the noncognitivist who does not base her position on the plausibility of Moore's original formulation of the OQA. If it is appropriate to determine the meaning of our moral terms by means of an updated causal theory of reference as the contemporary moral realist claims, then it is appropriate to determine the meaning of 'moral' itself through this same causal theory of reference. There is no basis for the moral realist to insist that the noncognitivist must determine the meaning of 'moral' on entirely a priori grounds. As a result of her failure to refute internalism - considered as an a posteriori generalization concerning the essential nature of actual moral discourse - the moral realist falls prey to the Mackie-style argument. It is not plausible to characterize actual moral language use as an essentially descriptive discursive activity because the referents of moral terms would have to be peculiar sorts of entities that are too queer to be believed in. The mere conceivability of the amoralist will not serve to deflect this a posteriori challenge.

2.14: Williams' Integrity Objection

In the light of the previous considerations, we can consider the plausibility of the externalist's response to a familiar normative problem that was first raised by Bernard Williams. If externalism requires a response to this problem that is unintuitive, we have additional grounds for rejecting

externalism. I will attempt to demonstrate that this is indeed the case. Williams describes two scenarios where hapless moral agents find themselves in circumstances where a significant degree of personal turmoil would accompany their acting in a manner that would be conducive to the realization of maximal aggregate utility. In the first scenario, George, a chemist, must work in a laboratory that conducts research into chemical and biological warfare in order to support his family as well as to prevent a more zealous colleague from taking the job, thereby speeding up the pace of this research. In the second scenario, Jim, a botanist, must personally execute an innocent man in order to prevent the imminent execution of nineteen others. According to Williams, these scenarios illustrate an important problem with utilitarianism.

A feature of utilitarianism is that it cuts out a kind of consideration which for some others makes a difference to what they feel about such cases: a consideration involving the idea, as we might first and very simply put it, that each of us is specially responsible for what he does, rather than for what other people do. This is an idea closely connected with the value of integrity. It is often suspected that utilitarianism, at least in its direct forms, makes integrity as a value more or less unintelligible.²⁴

As Samuel Scheffler notes, one of the notorious problems with Williams' seminal discussion is its unclarity at a crucial point.

Despite the suggestiveness of Williams's remarks, his formulation leaves obscure precisely what feature of

²⁴Williams, 1973, p.99.

utilitarianism is supposed to alienate the agent from his actions and hence to undermine his integrity.²⁵

Scheffler offers the following reformulation of Williams' objection which focuses on the complaint that the requirements of utilitarianism are unreasonably demanding for autonomous agents with important personal commitments. For the purposes of our discussion, we can assume that Scheffler's characterization of the difficulty here is accurate. In his words:

utilitarianism incorporates a conception of the right which requires each agent in all cases to produce the best available outcome overall. It requires the agent to pursue his projects, commitments, and personal relationships whenever and to the extent that doing so would have the best overall outcome impersonally judged, and to neglect or abandon them whenever and to the extent that **that** would have the best overall outcome impersonally judged. Utilitarianism thus requires the agent to allocate energy and attention to the projects and people he cares most about **in strict proportion** to the value from an impersonal standpoint of his doing so, even though people typically acquire and care about their commitments quite independently of, and out of proportion to, the value that their having and caring about them is assigned in an impersonal ranking of overall states of affairs. It is **this** feature of utilitarianism which may be thought to alienate the agent 'from his actions and the source of his action in his own convictions', and thereby to undermine his integrity.²⁶

According to David Brink, the preceding problem does not constitute a legitimate criticism of utilitarianism. He claims that utilitarianism is unable to account for the personal point of view only if it is regarded as a 'decision procedure'. If it is regarded only as a standard or criterion

²⁵Scheffler, 1982, p.8.

²⁶Scheffler, 1982, p.9.

of rightness, it is not open to this line of criticism. This is precisely how the externalist moral realist does regard utilitarianism. For her, utilitarianism is simply a classificatory scheme which specifies that those complex natural states of affairs which constitute the maximal potential aggregate utility available in various situations make up the reference class which is rigidly designated by moral terms like 'good' and 'right'. Moral language use is simply a variety of descriptive discourse which pertains to this particular class of states of affairs. In Brink's words:

The personal point of view is important, and we can recognize this without making morality capture its importance. The worries that the importance of the personal point of view raises can be viewed not as **moral worries** but as **worries about morality**. Our sympathies for Jim need not be moral sympathies. ... We may come to wonder whether Jim has reason or enough reason to do as morality requires in these circumstances. These are worries about the justification or supremacy of moral demands, not about the correctness of a utilitarian account of morality.²⁷

These considerations have an important effect on the way that Williams' 'integrity' objection to utilitarianism is understood by the externalist. According to the externalist, Williams' objection does not impugn the plausibility of utilitarianism as a moral theory. Rather, it simply impugns the rationality of moral requirements. As Brink claims,

the real worry latent in the conflict between utilitarianism and the personal point of view is not the worry **within** morality about which moral theory is correct but the worry **about** morality concerning the rationality or supremacy of moral demands. We can recognize the

²⁷Brink, 1989, p.278.

importance of the personal point of view without granting it **moral** importance.

Indeed, not only **can** we recognize the importance of the personal point of view without granting it moral importance; we **should not** concede it moral importance.²⁸

At another point, Brink considers a passage where Williams criticizes utilitarianism, Kantianism, and, indeed, any form of 'impartial morality'. Here, he claims that:

The fact that Williams here takes Kantian views of morality to be equally guilty of making unreasonable demands supports the claim that the real worry latent in the conflict between utilitarianism and the personal point of view is not the worry **within** morality about which moral theory is correct but the worry **about** morality concerning the rationality or supremacy of moral demands. We can recognize the importance of the personal point of view without granting it **moral** importance.²⁹

This externalist interpretation of the integrity objection can be contrasted with the noncognitivist's interpretation. According to my noncognitivist characterization of the nature of moral language use, Williams' refusal to accept 'impartial morality' because it makes unreasonable demands on autonomous agents is construed as an argument **within** morality over whether or not any form of impartial morality is a plausible moral theory. In making his argument, Williams presents two concrete examples with the intention of evoking certain intuitive responses which accrue from the UMP embraced by the reader. He is banking on the reader's being intuitively struck that it is not defensible to maintain that George and Jim are bad if they do not act in a

²⁸Brink, 1986, p.435.

²⁹Brink, 1989, p.280.

manner which is conducive to the realization of maximal aggregate utility in their respective scenarios. His purpose is to convince the reader that there is something wrong with utilitarianism since the FOMS it calls for in these situations are not defensible. The success of this strategy is dependent on the character of the UMP embraced by the reader. If the reader holds an UMP where the FOMS 'George's not taking the job is bad' and 'Jim's not shooting an Indian is bad' are defensible, these scenarios will not strike her as revealing a bothersome structural problem which affects the plausibility of utilitarianism. On the other hand, if the reader holds an UMP where these FOMS are not defensible, these scenarios will strike him as revealing a bothersome structural problem which affects the plausibility of utilitarianism. It is to this latter class of readers that Williams is appealing. He is attempting to show them that utilitarianism is an unacceptable moral theory because it holds certain FOMS, which they do not take to be defensible, to be defensible. That is to say, he is trying to convince them that the requirements of utilitarianism are not a plausible set of moral requirements. This seems to be a much more natural way of characterizing the problem that Williams is getting at here. It falls in line with the way that the typical actual moral language user is apt to construe the problem. Brink rejects the possibility of this type of characterization, however, because it is based on

an implausible semantic assumption. Namely, that according to this kind of approach:

It is simply part of the concept of a moral consideration that moral considerations necessarily provide reason or conclusive reason for action (cf. Chapter 3). But as we have seen, there is reason to reject this internalist assumption. The amoralist's challenge to the rationality of moral demands is not only intelligible but deserves to be taken seriously.³⁰

As we have seen, however, the appropriate response to the intelligibility of the amoralist is 'So What?'. The noncognitivist interpretation of the integrity objection cannot be disposed of as easily as Brink supposes.

The question remains, then, which of the two preceding interpretations is to be preferred? It is perhaps worthwhile to pause here and emphasize that, at this point, the plausibility of utilitarianism as a normative doctrine is not the issue at hand. I have brought up this particular discussion of utilitarianism in order to consider whether or not externalism is a plausible metaethical doctrine. In order to answer this question, we should keep in mind that what is at issue here centers around what we take to be the correct characterization of the meaning of the word 'moral'. Externalists interpret the difficulty that Williams raises in a particular way because they define the word 'moral' in a certain way. For the externalist moral realist, moral language use consists of the application of a range of predicates like 'good' to fairly well-defined types of states of affairs.

³⁰Brink, 1989, pp.278,279.

According to the utilitarian externalist moral realist, the predicate 'good' is correctly applied to those persons, actions, and states of affairs whose existence is conducive to or constitutive of the realization of maximal aggregate utility. For her, Williams, properly understood, is simply suggesting that autonomous agents will, on occasion, have no good reason to act in a manner that is conducive to the realization of maximal aggregate utility. This is a complaint against the rationality of morality, not a complaint against the plausibility of utilitarianism as a moral theory. The non-rationality of a particular requirement does not negate its status as a bona fide moral requirement.

2.15: A Problem For The Externalist

It seems to me that the externalist moral realist's response to Williams 'integrity' objection can actually be used as a reductio against her position. The externalist moral realist is committed to this kind of response to Williams' integrity objection and there are two good reasons for rejecting such a response. The first reason centers on the characterization of moral discourse presented in this chapter. Moral language use is an intrinsically communal discursive activity whose essential nature can only be determined a posteriori by consulting the practices and linguistic intuitions of actual moral language users. In practice, people do take the acceptance that 'x is bad' to constitute the acceptance of a reason not to do x. To hear someone say 'I

admit that *x* is bad, but I don't see what that has to do with whether or not I should *x*' will strike the normal moral language user as odd. It is an a posteriori fact that moral language is **not** used in this way. The externalist interpretation of Williams' integrity objection presupposes the legitimacy of a characterization of moral language use that **actual** moral language users would take to be very odd.

Such an interpretation of the nature of moral discourse can be shown to be deviant by considering the implications of treating this characterization of moral discourse as if it were the norm. If externalism is correct, there is no principled reason why we cannot do this. Would the 'moral' discourse engaged in by an entire community of externalists strike actual moral language users as the same kind of discursive activity that they engage in? It would seem not, for the same reason that the discursive activity engaged in by an entire community of amoralists would not strike actual moral language users as the same kind of discursive activity they are engaged in. It lacks an **essential** feature of actual moral discourse. In actual moral language use, moral attributions **are** taken to imply reasons for action. This explains why moral attributions can function as criticisms in actual moral discourse.

If the externalist's characterization of the nature of moral language use is correct, however, the mere recognition of a moral requirement is never in itself sufficient grounds

for commanding our allegiance. This principle does not simply apply to certain moral requirements on certain occasions. It applies to any moral requirement on any occasion. According to the externalist's scheme, only certain moral requirements command our allegiance and these do so on grounds which are independent of their being moral requirements. Thus, the externalist is committed to the claim that there is no conceptual connection at all between an action's being something we should do and that action's being a moral requirement. As Brink freely admits

This interpretation of the conflict between utilitarianism and the personal point of view makes the externalist assumption that it is a substantive question whether moral considerations provide an agent with reason or sufficient reason for action, the answer to which depends on a substantive theory of reasons for action or rationality and, in all probability, a substantive theory of human welfare.³¹

For any particular moral requirement, the answer to this substantive question is going to depend on the nature of the independent grounds which can be given for complying with that moral requirement. For the externalist moral realist, then, there will be a certain range of moral considerations which **do** provide agents with reason or sufficient reason for action, and a certain range of moral considerations which **don't** provide agents with reason or sufficient reason for action. Recall that, for the utilitarian externalist moral realist, any action that is conducive to the realization of maximal

³¹Brink, 1986, p.413.

aggregate utility has, by that very token, a moral consideration in its favor. Of those moral considerations which **do** there must be something about them over and above their merely being moral considerations which distinguishes them from the range of moral considerations which **don't**. It is the presence of this additional feature which provides agents with reason for action in the former range of cases but not in the latter. In order to criticize anyone's actions, it will be necessary to explicitly refer to this additional feature. It will not be enough to simply point out that her actions transgress the requirements of morality

If this characterization of the nature of moral language use is correct, then the claim that a particular action is morally required could **never**, by itself, function as a reason to act in that manner. As a result, merely pointing to someone's failure to conform to the requirements of morality could never serve, in itself, as a means of criticizing her actions. According to the externalist position, whether or not a person could be legitimately criticized for acting in a particular manner is an issue which is quite independent of whether or not his behavior conforms to the requirements of morality. In certain instances, it might be quite reasonable for that person to resist the requirements of morality. In such cases, she could not be legitimately criticized for transgressing the requirements of morality. Thus, in order to criticize anyone's actions, it would always be necessary to

point to specific features of those actions above and beyond their being wrong or bad things to do. Otherwise, your comments about their actions' being wrong or a bad thing to do would not function as a criticism. Neither would they be taken as a criticism of your actions.

Yet, the preceding characterization is not an accurate reflection of the nature of actual moral discourse. In actual moral discourse, people do take the claim that an action is morally required to serve as a behavioral demand. For instance, telling someone that drinking alcohol is morally wrong functions, all by itself, as a criticism of that action. It is not necessary to point to an additional feature of consuming alcohol in order to indicate that this action is being censured. Further, typical moral language users who drink alcohol never attempt to defend this action by assenting to its moral wrongness while claiming that it is a special type of moral requirement which they have no good reason to conform to. Typical moral language users never attempt to defend **any** action by agreeing that the action is morally wrong but claiming that it is a special type of moral requirement which they have no good reason to conform to. Rather, they defend their actions by **denying** that they are morally required to refrain from acting in such a manner. In short, typical moral discourse is simply not conducted along externalist lines. This is an important empirical fact. Since externalism does not provide an accurate a posteriori characterization of

the nature of actual moral language use, it is unacceptable on those grounds. Since this characterization of the nature of moral discourse is entailed by the truth of externalism, externalism can be rejected on these a posteriori grounds.

2.16: A Second Problem For The Externalist

The second reason that the externalist's interpretation of Williams' integrity objection is not plausible centers on the Moral Twin-Earth argument presented in the previous chapter. If we suppose, for the moment, that a utilitarian version of externalist moral realism is correct, then **moral** requirements are those which are conducive to the realization of those states of affairs which instantiate maximal aggregate utility. If this is so, then the meaning of 'moral' is already taken, so to speak. Requirements which are based on sets of criteria other than conduciveness to the realization of maximal aggregate utility, are, by that very token **not moral requirements**. If we suppose that Kantian requirements will lead to distinctly different results from adherence to utilitarian requirements in a significant range of situations, Kantian requirements must not be moral, but, say, '**schmoral**' requirements. The problem does not stop at Kant. The requirements of each distinct position on the normative spectrum will have its own peculiar set of requirements. An ethics of caring will have '**spmoral**' requirements. An Aristotelian ethics will have '**skmoral**' requirements. A Nietzschean ethics will have '**stmoral**' requirements. The list

can go on indefinitely. Again, taking the externalist line seems to result in a highly anti-intuitive consequence. What are we to make of this regress?

According to Robert Bright, that's just the way it is. The only people who would be bothered by such a regress are those who have a fundamentally mistaken concept of morality. There are two related mistakes to be considered here. The first

mistake in all this,..., is to suppose that a moral theory should be centrally preoccupied with right and wrong in the first place.³²

The second related

mistake, then, is to take the shortcut of conceiving of consequentialist theories as theories of obligation, and thereby bless deontological concepts with a moral standing they don't deserve.³³

If we don't begin by making these mistakes, Williams' integrity objection can be handled in a straightforward manner. In the second scenario that Williams describes, the problem Jim faces is simply whether or not to act morally. The problem does not center on whether or not he has an obligation to maximize utility in that situation by shooting one of the Indians. Thus, the Pedro scenario presents no problem for the plausibility of utilitarianism qua moral theory.

³²Bright, 1991, p.218. Although Bright's comments come from an unpublished doctoral dissertation, it seemed appropriate to include them in my discussion since they constitute a particularly stark representation of the externalist presuppositions I am opposing.

³³Bright, 1991, p.227.

To put a point on it, obligations seem more-or-less irrelevant to assessing Jim's case. What is relevant is that it is impossible to make sense of Jim's predicament, if we imagine ourselves in Jim's shoes, unless we credit Jim or ourselves with a clear recognition that one of the actions available in the circumstances is of substantially greater moral worth than the other, though substantially less attractive from a personal point of view.³⁴

According to this strategy, we can explain away the regress problem by mere stipulation. We can just decide that 'moral' is going to be defined in terms of utility. Once we have decided this, we can assess the moral worth of whatever we want, including deontological principles, in terms of the manner in which they measure up to the standard of utility. If standards aren't conducive to the realization of maximal utility, then they lack moral worth. Thus, they cannot be moral standards. As Bright claims:

The principle of utility is as capable of deciding the moral worth of various practices governing rewards and sanctions as it is of assessing the worth of actions generally, and hence deontological practices are morally criticizable like other practices. ... The principle of utility tells us what is morally better and worse for us to do, including that there are morally better and worse ways of imposing obligations on ourselves and others, but it does not and cannot directly impose obligations on us; to suppose that it could is a bit of superstitious nonsense ("nonsense on stilts").³⁵

Unfortunately, all we really have here is a typical instance of externalist semantic legislation. The crucial issue here centers around the correct determination of the meaning of the word 'moral'. Instead of providing reasons for defining

³⁴Bright, 1991, pp.235,236.

³⁵Bright, 1991, p.237.

'moral' in a particular way, Bright simply uses it in a way that favors his externalist perspective. He insists on couching the discussion in terms of certain actions and states of affairs being 'morally better' than others and refuses to countenance the possibility that the notion of 'moral obligation' is a legitimate concept. This seems especially strange since the typical moral language user is fairly at home with the notion of moral obligations while the idea of certain states of affairs being 'morally better' than others has a decidedly odd ring to it. To say that some states of affairs are **better** than others makes perfect sense to the normal moral language user, but one may doubt whether calling one state of affairs **morally better** than another does. In any case, since he begins his discussion by begging the crucial question here, his comments provide no principled defense for the externalist interpretation of the meaning of the word 'moral'. As a result, his comments do not alleviate the intuitive difficulties that the regress problem raises for actual moral language users.

As I have noted earlier, there are strong a posteriori grounds for rejecting the hard externalist line that leads to the regress problem. Unfortunately, or fortunately, as the case may be, typical moral language users are **not** utilitarian externalist moral realists. Typical moral language users do not count non-utilitarian considerations as **non-moral** considerations. It **might** have been the case that typical moral

language users were utilitarian externalist moral realists. If they were, they would count the requirements of non-utilitarian normative theories as non-moral requirements. They would also, as we have seen, find the Moral Twin-Earth argument completely unconvincing. But, as a matter of a posteriori fact, they don't. Hard line externalists have to interpret Williams' integrity objection in the way that they do, but, in doing so they merely betray the fact that they are out of step with the rest of the linguistic community in the way that they use the word 'moral'. As such, their pronouncements concerning the nature of morality should not be given undue weight. For the externalist, there will be as many different varieties of normative requirements, of which **moral** requirements can only be one, as there are different normative systems. As a result, we can reject externalism on the a posteriori grounds that it calls for a characterization of the essential nature of moral discourse that is intuitively unacceptable to actual moral language users. What they take to be yet another strength of their position actually functions as a *reductio* against it. Thus, we can reject the utilitarian externalist moral realist's interpretation of Williams' integrity objection to utilitarianism. Can we thus reject utilitarianism as an acceptable moral theory? The answer to that question depends on what one's UMP is. Is there a way we can decide amongst competing UMPs? I will give my answer to that question in the final chapter of the dissertation.

CHAPTER THREE: THE EMBEDDING PROBLEM

3.1: Introduction

Having provided a preliminary argument in favor of a noncognitivist construal of moral language in the first chapter, I went on to articulate more fully the nature of the version of noncognitivism I am espousing in the early part of the second chapter. From there, I explained why this updated version of noncognitivism is immune to the amoralist argument, one of the two principal criticisms of traditional noncognitivism. In the third chapter, I will continue along my defensive tack by attempting to explain why the version of noncognitivism I espouse is immune to the second principal criticism of traditional noncognitivism: the embedding problem. I will begin with a brief description of what is at issue here. The problem itself is fairly simple to understand. We commonly incorporate moral statements into our reasoning patterns in ways that seem to make sense only if these statements are cognitive. The noncognitivist does not seem to be in a position to be able to justify patterns of reasoning which are clearly justifiable. Simon Blackburn illustrates this difficulty with the following illustration which he borrows from Peter Geach's earlier discussion. I will use this simple inference as the template for the solution to the embedding problem which will be provided in the latter part of this chapter.

Blackburn's illustration is as follows:

- [1] It is wrong to tell lies.
- [2] If it is wrong to tell lies, it is wrong to get your little brother to tell lies.
- So [3] It is wrong to get your little brother to tell lies.

This is a valid argument, illustrating the general form: P; if P then Q; so Q. But the argument is only of this form because the sentence 'It is wrong to tell lies' means the same on each occurrence. ... The question now is: how does an expressive theory explain the identity of meaning? For anyone asserting the second, hypothetical premise is **not** expressing an attitude of condemnation towards telling lies. He commits himself to no attitude towards it at all.¹

The problem is determining how to treat moral expressions when they occur in these 'unasserted' or 'embedded' contexts. This problem does not arise for the moral realist, of course. For her, moral statements are straightforward cognitive statements whose **meaning** is not affected by being placed in an embedded context. If the embedding problem cannot be solved, it constitutes a reason for rejecting noncognitivism which is independent of the amoralist argument.

In a recent essay, Peter Railton makes the following comments on this issue:

One of the tasks facing expressivists is to provide a systematic reinterpretation of discourse about a person's good in order to show how essentially expressive, noncognitive language could display on its surface thoroughly descriptive, cognitive grammatical and logical behavior. Numerous technical problems beset such reinterpretations, and no one has as yet offered an expressivist account that satisfactorily preserves all the cognitive features of discourse about value.²

¹Blackburn, 1984, p.190.

²Railton, 1989, p.153.

This is not to say that moral realism is not without its problems. According to Railton, in the debate between naturalism and expressivism (or what I have been calling realism and noncognitivism), each camp faces a difficulty which is trivial to the other side.

Just as an expressivist faces the problem of showing how essentially noncognitive uses of language could have the cognitive character of actual value discourse, the naturalist cognitivist faces the complementary problem of explaining how an essentially descriptive use of language could have the prescriptive force of value discourse.³

3.2: Geach And 'The Frege Point'

In order to understand how to address the embedding problem adequately, it will be best to begin by considering its historical origin. One of its earliest formulations occurs in an article by Peter Geach in 1965 entitled 'Assertion'. In this article, Geach focused on fleshing out the logical implications of a semantic doctrine which he refers to as 'The Frege Point'. The Frege point is that

A thought may have just the same content whether you assent to its truth or not; a proposition may occur in discourse now asserted, now unasserted, and yet be recognizably the same proposition.⁴

The important logical implication of The Frege Point is that its truth is required in order to demonstrate the validity of even patently valid inference patterns like modus ponens and disjunctive syllogism. The reason that the truth of The Frege Point is required is quite simple. If The Frege Point were not

³Railton, 1989, p.154.

⁴Geach, 1965, p.449.

true, we would have no grounds for maintaining that every instance of a particular atomic sentence that occurs in an argument stands for the same proposition. As a result, we would have no basis for assigning the same truth value to every occurrence of that atomic sentence. Thus, we could not provide truth-functional accounts of valid inference patterns that are based on the truth values of their constituent complex sentences. In order to provide such an account, it must be the case that the atomic constituents of these complex sentences are assigned truth values on a basis which is entirely independent of whether or not such atomic sentences are actually asserted. This is just to say that The Frege Point must be true.

This presents a prima facie difficulty for the plausibility of noncognitivism because it is not immediately clear how a noncognitivist characterization of moral language can satisfy The Frege Point. Geach was well aware of this. He made the following comments about 'anti-descriptive' theories like noncognitivism:

Predicates of a philosophically exciting sort have been badly misconstrued because assertoric force has been supposed to inhere in them. Theory after theory has been put forward to the effect that predicating some term "P" - which is always taken to mean: predicating "P" assertorically - is not describing an object as being P but some other "performance"; and the contrary view is labelled "the Descriptive Fallacy." All these theories are constructed on the same pattern and admit, as we shall see, of the same refutation.⁵

⁵Geach, 1965, p.461.

The problem lies in the theory of meaning that such anti-descriptivist theories are essentially committed to. For our purposes, we can confine ourselves to the case of the noncognitivist's construal of moral terminology. Geach uses modus ponens to illustrate the difficulty faced by the noncognitivist here. He begins by noting that:

Of course, the anti-descriptive theorist will reply that his theory was not meant to cover such cases - that the same form of words, after all, may have different uses on different occasions. This possibility of varying use, however, cannot be appealed to in cases where an ostensibly assertoric utterance "p" and "If p, then q" can be teamed up as premises for a **modus ponens**. Here, the two occurrences of "p", by itself and in the "if" clause, must have the same sense if the **modus ponens** is not to be vitiated by equivocation; and if any theorist alleges that at its ostensibly assertoric occurrence "p" is really no proposition at all, it is up to him to give an account of the role of "p" that will allow of its standing as a premise.⁶

He then offers the following example of a modus ponens inference consisting of atomic sentences which contain the moral predicate 'bad'.

If doing a thing is bad, getting your little brother to do it is bad.

Tormenting the cat is bad.

Ergo, getting your little brother to torment the cat is bad.

The whole nerve of the reasoning is that "bad" should mean exactly the same on all four occurrences - should not, for example, shift from an evaluative to a descriptive or conventional or inverted commas use.⁷

⁶Geach, 1965, p.463.

⁷Geach, 1965, pp.463,464.

3.3: Searle's Parallel Criticism

Three years earlier, John Searle advanced a similar objection to the plausibility of noncognitivism. According to Searle, noncognitivists, like all anti-descriptivists, have engaged in a fundamental theoretical error which can be described in the following manner:

the statement that a word is used to perform certain kinds of speech acts is taken by philosophers who say this sort of thing to be a statement about the meaning of the word and hence to be part of a philosophical analysis or explication of the word.⁸

For example, the meaning of 'good' lies in its commendatory function while the meaning of the word 'bad' lies in its condemnatory function. The problem is that when words like 'good' and 'bad' occur in certain grammatical contexts such as the antecedent of a conditional statement, they are clearly not being used to perform their characteristic evaluative function. Thus, if noncognitivism is taken to be providing an analysis of the meaning of these words, it clearly fails when these words occur in embedded contexts. The initial plausibility of a noncognitivist analysis of moral terms is based on a fundamental mistake. In Searle's words:

the mistake is to suppose that an analysis of calling something good gives us an analysis of "good". This is a mistake because any analysis of "good" must allow for the fact that the word makes the same contribution to different speech acts, not all of which will be instances of calling something good. "Good" means the same whether I ask if something is good, hypothesize that it is good, or just assert that it is good. But only in the last does

⁸Searle, 1962, p.423.

it (can it) have what has been called its commendatory function.⁹

As a result, we have grounds for claiming that noncognitivism, by its very nature, is unable to provide adequate analyses of the meanings of moral terms. Together, the parallel criticisms of Geach and Searle form the historical basis for what has come to be known as 'The Embedding Problem'.

3.4: The Target: R.M. Hare's Noncognitivism

It is important to notice that both of these criticisms were principally and explicitly directed towards the noncognitivism espoused by R.M. Hare. This is not surprising since Hare's version of noncognitivism was clearly the most influential of the time. In order to clarify the essence of the problem that Geach and Searle were focusing on, it will be useful to describe the moral semantics of R.M. Hare in more detail. According to Hare, the meaning of the word 'good' is essentially tied to its function of **commending**. This commending function can be used to explain why moral vocabulary cannot be defined in terms of natural properties. According to Hare:

we cannot say that 'x is a good A' means the same as 'x is an A which is C', because then it becomes impossible to commend A's which are C by saying 'A's which are C are good A's'.¹⁰

Hare's argument here does not rely on any contingent facts about the way that actual moral language users **do** use moral

⁹Searle, 1962, p.429.

¹⁰Hare, 1952, p.89.

predicates. If it did, it would be an a posteriori argument in favor of characterizing the meaning of moral predicates in a certain way. Instead, his argument focuses on eliminating the possibility of providing a suitable **analysis** of the word 'good' in terms of natural properties. The reason it is impossible to commend A's for being C in this situation is that 'C-ness' and 'goodness' are **synonymous**. Thus, we have an **a priori** case for the claim that the meaning of 'good' must be cashed out in terms of its characteristic commending function rather than in terms of any cluster of natural properties which might typically be deemed to be good. The reasoning involved here parallels Moore's reasoning in the original OQA. In fact, Hare regarded the preceding piece of reasoning as an updated version of Moore's OQA. He described Moore's position as follows:

Moore thought that he could prove that there were no such defining characteristics for the word 'good' as used in morals. His argument has been assailed since he propounded it; and it is certainly true that the formulation of it was at fault. But it seems to me that Moore's argument was not merely plausible; it rests, albeit insecurely, upon a secure foundation; there is indeed something about the way in which, and the purposes for which, we use the word 'good' which makes it impossible to hold the sort of position Moore was attacking, although Moore did not see clearly what this something was. Let us, therefore, try to restate Moore's argument in a way which makes it clear why 'naturalism' is untenable, not only for the moral use of 'good' as he thought, but also for many other uses.¹¹

Hare then argued that the crucial difficulty with any proposed naturalistic definition of 'good' is that it would preclude

¹¹Hare, 1952, pp.83,84.

the crucial commending function which is the earmark of evaluative vocabulary.

Let us generalize. If 'P is a good picture' is held to mean the same as 'P is a picture and P is C', then it will become impossible to commend pictures for being C; it will be possible only to say that they are C. It is important to realize that this difficulty has nothing to do with the particular example I have chosen. It is not because we have chosen the wrong defining characteristics; it is because, whatever defining characteristics we choose, this objection arises, that we can no longer commend an object for possessing those characteristics.¹²

Hare went on to claim that although words like 'good' have both a descriptive and an evaluative aspect to their meaning, the evaluative aspect is primary. There are two reasons for this:

First, the evaluative meaning is constant for every class of object for which the word is used. When we call a motor-car or a chronometer or a cricket-bat or a picture good, we are commending all of them. But because we are commending all of them for different reasons, the descriptive meaning is different in all cases. ...

The second reason for calling the evaluative meaning primary is, that we can use the evaluative force of the word in order to **change** the descriptive meaning for any class of objects.¹³

Hare's a priori analysis of the meaning of moral terminology falls into trouble, however, when we consider the function of moral vocabulary when it occurs in embedded contexts. Since tokens of moral vocabulary aren't being used to evaluate in these contexts, it would appear that they have lost the primary aspect of their meaning when they occur there. Such

¹²Hare, 1952, p.85.

¹³Hare, 1952, pp.118,119.

occurrences of moral terms fail to retain the essential element which makes them distinctly moral vocabulary. Thus, Hare has no grounds for maintaining that such gelded tokens of moral vocabulary have the same content as their virile counterparts that do not occur in embedded contexts. He has already committed himself to the claim that we can know *a priori* that genuine instances of moral vocabulary can be identified by their distinctive evaluative function. Since moral terms in embedded contexts do not perform the required evaluative function, they cannot be instances of genuine moral terminology. They do not satisfy the evaluative criterion which is the earmark of all genuine moral terminology.

3.5: Hare's Dilemma

The preceding characterization of the meaning of moral vocabulary leaves Hare in an uncomfortable dilemma, however. On the one hand, he is committed to maintaining that evaluative meaning is the crucial feature which distinguishes moral from non-moral terminology. As we saw earlier, this furnishes him with an analytic basis for maintaining that moral discourse is fundamentally prescriptive rather than descriptive. Since the basis for his position is his analysis of the word 'good', he has *a priori* grounds for espousing noncognitivism. If he does this, however, he is at a loss to explain how obviously legitimate inference patterns containing moral terminology can be shown to be legitimate. Since noncognitivism requires that an essential element of the



meaning of a moral utterance consists in its being asserted, a noncognitivist like Hare has no grounds for maintaining that an atomic sentence which contains a moral term and is asserted means the same thing as an identical token of that atomic sentence which isn't asserted. According to the noncognitivist's own criterion of meaning, an atomic sentence that occurs in an unasserted context cannot mean the same thing as an identical token that is asserted. If this is so, however, the noncognitivist's characterization of the nature of moral language use no longer possesses the means for demonstrating the validity of inference patterns which are patently valid. It loses this ability because it is unable to maintain that all the tokens of a particular atomic sentence which contains moral terminology have the same content. The only way to make such inferences appear to be valid is by **equivocating** between the embedded and non-embedded occurrences, since, by the noncognitivist's own criterion of meaning, they cannot mean the same thing. Thus, every time we make inferences which involve moral terminology, we are engaging in a subtle logical cheat. We are treating groundless inferences as if they were valid. But this position is fraught with intuitive difficulty. It is simply not plausible to maintain that we engage in subtle fallacies of equivocation every time we make inferences involving straightforward moral statements and their embedded counterparts. There just doesn't seem to be anything untoward about making inferences which

involve moral statements. Even noncognitivists who are made aware of the equivocation problem are unwilling to reject otherwise valid inference patterns whenever they contain moral terminology. The fact is that moral statements fit quite naturally into our common patterns of inference.

On the other hand, the noncognitivist doesn't seem to be in a position to be able to admit that an inference pattern which is valid when it only contains descriptive statements is still valid even when it contains moral statements. If that inference is still valid, then it cannot involve subtle equivocations. But if it doesn't involve subtle equivocations, then the embedded and non-embedded tokens of any particular moral term it contains must mean the same thing. But if these embedded and non-embedded tokens must mean the same thing, then the primary meaning of moral terms cannot lie in their evaluative function. But if the primary meaning of moral terms does not lie in their evaluative function, then noncognitivism must be false since noncognitivism just is the claim that the primary meaning of moral terms lies in their evaluative function.

3.6: Noncognitivism As An A Posteriori Doctrine

At this point, it is important to emphasize once more that I am not interested in providing a defense of all versions of noncognitivism which have been historically espoused. Rather, I am interested in providing a defense of the particular version of noncognitivism I have been

describing. I am not interested in defending R.M. Hare.¹⁴ Though he had his own misgivings about Moore's precise formulation of it, Hare was still convinced that Moore's OQA showed that it is not possible to provide a naturalistic analysis of the meaning of moral terminology. He still agreed with the linguistic presupposition that since a moral term like 'good' is not synonymous with any term that refers to a natural property or type of natural state of affairs, its meaning cannot be analyzed in terms of any natural property or type of natural state of affairs. But like other classical noncognitivists, Hare took Moore's OQA in a new direction. As Putnam points out:

His defense of non-naturalism backfired; his students may have been convinced by Moore that there were such things as 'non-natural properties' (although Russell, at least, was to lose the faith) but later philosophers of a naturalistic kind tended to feel that Moore had provided a **reductio ad absurdum** of the idea that there are such things as value properties.¹⁵

Thus, armed with their pre-Kripke-Putnam intuitions about the nature of meaning, Hare and other classical noncognitivists set out to provide alternative a priori analyses of moral language which cashed out the meaning of moral terminology in terms of an evaluative function. It is important to realize that the parallel criticisms of Geach and Searle were intended

¹⁴This is not to say that Hare had nothing to say on his own behalf. Rather than attempting to solve the embedding problem, he attempted to formulate a 'logic of moral concepts'. See Hare, 1963, Ch.10 - Logic And Morals.

¹⁵Putnam, 1981, p.206.

to be applied against classical noncognitivism. Since the version of noncognitivism I am espousing does not purport to be providing an a priori analysis of the meaning of moral terminology, the fact that FOMS often occur in embedded contexts does not present an immediate problem for it. It is not incumbent upon me to demonstrate how earlier versions of noncognitivism might deal with the embedding problem. A crucial difference between my version of noncognitivism and classical versions like Hare's is that my version does not purport to provide an a priori analysis of the meaning of moral terms. Rather, it purports to be providing an accurate a posteriori characterization of the nature of actual moral language use. As such, it must account for the a posteriori fact that we find it quite natural to employ accepted inference patterns which contain FOMS. Furthermore, it must account for this fact in a way that treats such FOMS as genuinely noncognitive. If I am able to do this, I will have shown that the embedding problem does not pose a difficulty for the position I am defending. This is all that is required of me.

3.7: Gibbard's Solution To The Embedding Problem

In order to accomplish this, I will make use of the work of Allan Gibbard. He provides a solution to the embedding problem which is intended to apply to the particular version of noncognitivism he espouses. He calls this version 'norm expressivism'. Insightful as it is, it seems to me that

Gibbard's solution to the embedding problem is unnecessarily cumbersome. A twofold task remains, then. First, I will describe Gibbard's solution to the embedding problem. Next, I will explain how to formulate an alternative solution to the embedding problem that makes use of some of Gibbard's insights.

Gibbard begins by offering the following characterization of normative semantics:

A normative sentence, the expressivist says, expresses a state of mind; its meaning is explained not by giving truth conditions but by telling what state of mind it expresses.¹⁶

An important point to notice here is that Gibbard treats normative **sentences**, rather than normative **predicates**, as the fundamental units of meaning. Normative sentences are the fundamental units of meaning because they correspond to particular states of mind, or intentional states, while normative predicates do not. Rather than attempting to explain how type-identical normative predicates can mean the same thing in embedded and non-embedded contexts, the focus is on explaining how type-identical normative sentences can have the same content in both embedded and non-embedded contexts. The problem for the expressivist then, is to explain how a normative sentence which occurs in an embedded context refers to the same state of mind that it expresses when it occurs in a non-embedded context. The first step to providing an answer

¹⁶Gibbard, 1990, p.92.

to this question is to ascertain the content of this state of mind. As a first approximation of the state of mind in question, Gibbard considers the notion of 'normative content'. "The normative content of a statement is the set of fully opinionated states one could be in and still accept the statement."¹⁷ Each of these fully-opinionated states consists of two components: a factual one and a normative one. The factual one pertains to the complete set of beliefs the agent has about the way that the world is. The normative one pertains to the complete set of general norms which the agent espouses. We can represent the vast array of different possible combinations of these two components with a series of ordered pairs $\langle w, n \rangle$. Each of these ordered pairs will stand for a different 'factual-normative world'. "Any particular normative judgment **holds**, as a matter of logic, in the factual-normative world $\langle w, n \rangle$."¹⁸ The content of a particular normative judgment then, can be represented by the set of all factual-normative worlds in which it holds.

In order to illustrate the point Gibbard is making here, we can consider a simple matrix consisting of the nine factual-normative worlds which result from the combination of three possible worlds with three normative positions. The relevant aspects of the three possible worlds are as follows:

¹⁷Gibbard, 1990, p.94.

¹⁸Gibbard, 1990, p.95.

w_1 - He hasn't stolen from you, but if you steal from him, he will never miss it and the theft will greatly benefit you.

w_2 - He has stolen from you, and if you steal from him, he will never miss it and the theft will greatly benefit you.

w_3 - He has stolen from you, but if you steal from him, the theft will cause him more grief than you good.

The relevant aspects of the three different normative positions are as follows:

n_1 - Reciprocation: Steal from others if, and only if, they have stolen from you.

n_2 - Crude Act Utilitarianism: Steal from others when, and only when, it will increase immediate global utility.

n_3 - Kantianism: Never steal from anyone under any circumstances.

The resulting matrix will look like this:

$\langle w_1, n_1 \rangle,$	$\langle w_1, n_2 \rangle,$	$\langle w_1, n_3 \rangle,$...
$\langle w_2, n_1 \rangle,$	$\langle w_2, n_2 \rangle,$	$\langle w_2, n_3 \rangle,$...
$\langle w_3, n_1 \rangle,$	$\langle w_3, n_2 \rangle,$	$\langle w_3, n_3 \rangle,$...
\vdots	\vdots	\vdots	

The dots on the bottom and right sides of the matrix indicate that the limited set of factual-normative worlds considered in this simple illustration can be infinitely expanded. We can now apply the following normative statement, which we will call 'S', to the matrix: 'Stealing from him on this occasion

is bad.' S holds in the following set of factual-normative worlds:

$\langle w_1, n_1 \rangle, \langle w_1, n_3 \rangle, \langle w_2, n_3 \rangle, \langle w_3, n_2 \rangle, \langle w_3, n_3 \rangle, \dots$

We can call this set O_1 . S does not hold in the remaining factual-normative worlds:

$\langle w_1, n_2 \rangle, \langle w_2, n_1 \rangle, \langle w_2, n_2 \rangle, \langle w_3, n_1 \rangle, \dots$

We can call this set \bar{O}_1 . Anyone who sincerely avows S will be in a credal-normative state corresponding to one of the factual-normative worlds in O_1 . Anyone who sincerely disavows S will be in a credal-normative state corresponding to one of the factual-normative worlds in \bar{O}_1 . Thus, if someone espouses S, her state of mind cannot be characterized by any of the factual-normative worlds in \bar{O}_1 . It must correspond to one of the factual-normative worlds in O_1 . Linking normative statements to fully-defined sets of ordered pairs provides us with a semantic basis for resolving the embedding problem. In Gibbard's system, it allows us to affix each non-truth-functional normative statement to a statement which is truth-functional. In order to do this, we simply

Replace each normative predicate in S with its n-corresponding descriptive predicate. That yields a purely descriptive statement S_n . Then normative statement S holds in $\langle w, n \rangle$ if and only if S_n holds in w.¹⁹

We can now use Blackburn's simple modus ponens example to illustrate Gibbard's solution to the embedding problem.

¹⁹Gibbard, 1990, p.96.

Gibbard's framework construes the inference in the following manner:

(1a) 'It is wrong to tell lies' holds in $\langle w, n \rangle$, for all $\langle w, n \rangle$ in O_L .

(2a) 'If it is wrong to tell lies, it is wrong to get your little brother to tell lies' holds in $\langle w, n \rangle$, for all $\langle w, n \rangle$ in O_C .

(3a) 'It is wrong to get your little brother to tell lies' holds in $\langle w, n \rangle$, for all $\langle w, n \rangle$ in O_B .

O_L stands for the set of factual-normative worlds within which 'It is wrong to tell lies' holds. O_C stands for the set of factual-normative worlds within which 'If it is wrong to tell lies, it is wrong to get your little brother to tell lies' holds. O_B stands for the set of factual-normative worlds within which 'It is wrong to get your little brother to tell lies' holds. Replacing the normative predicate 'wrong' with the descriptive predicate 'n-forbidden' we get:

(1b) 'It is n-forbidden to tell lies' holds in w , for all $\langle w, n \rangle$ in O_L .

(2b) 'If it is n-forbidden to tell lies, it is n-forbidden to get your little brother to tell lies' holds in w , for all $\langle w, n \rangle$ in O_C .

(3b) 'It is n-forbidden to get your little brother to tell lies' holds in w , for all $\langle w, n \rangle$ in O_B .

This reformulation of the argument allows us to employ the standard criterion for determining validity: If an inference

is valid, then if all the premises are true, the conclusion must be true as well. In Gibbard's words:

Embedded contexts now pose no special problem: so long as we have an account of what it is for a descriptive statement to hold in a possible world w - for a maximally specific way the world might be - we can now say what it is for any normative statement to hold for a factual-normative world $\langle w, n \rangle$.²⁰

Of course, we do have an account of what it is for a descriptive statement to hold in a possible world. A descriptive statement holds in a possible world if it is true in that world. This inference is valid because it is not possible to have a world w where (1b) and (2b) are true and (3b) is false. Since this holds for any normative system n , it is not possible to be in a credal-normative state $\langle w, n \rangle$ where (1a) and (2a) hold and (3a) doesn't. So, the inference of (3) from (1) and (2) must also be valid. Thus, Gibbard's expressivist semantics ultimately relies on truth conditions. He justifies inference patterns involving noncognitive expressions which are not truth-functional by piggybacking them upon a co-extensive set of descriptive statements which are truth-functional. This allows the inference patterns containing noncognitive expressions to inherit a kind of 'validity by association'.

3.8: Elaborating On Some Of Gibbard's Insights

Intriguing as Gibbard's solution is, it seems to me that it is not necessary for the noncognitivist to take this

²⁰Gibbard, 1990, p.97.

circuitous route. The problem with Gibbard's indirect solution lies in the implicit assumption that any suitable semantic undergirding for a system of inference rules must be couched in terms of truth conditions. Since noncognitive utterances cannot be truth-functional by their very nature, he is forced into justifying inferences involving noncognitive expressions by an indirect piggybacking strategy. If we do not accept the assumption that any suitable semantic undergirding for a system of inference rules must be couched in terms of truth conditions, however, we are not forced into taking the indirect route.

I will now describe a direct solution to the embedding problem which avails itself of some of Gibbard's insights. We can begin by considering a parallel version of the simple matrix considered earlier. This matrix will consist of nine **factual-moral worlds** which result from the combination of three possible worlds with (and here is the difference) three UMPs. The parallel matrix will look like this:

$$\begin{array}{lll}
 \langle w_1, \text{UMP}_1 \rangle, & \langle w_1, \text{UMP}_2 \rangle, & \langle w_1, \text{UMP}_3 \rangle, \dots \\
 \langle w_2, \text{UMP}_1 \rangle, & \langle w_2, \text{UMP}_2 \rangle, & \langle w_2, \text{UMP}_3 \rangle, \dots \\
 \langle w_3, \text{UMP}_1 \rangle, & \langle w_3, \text{UMP}_2 \rangle, & \langle w_3, \text{UMP}_3 \rangle, \dots \\
 \vdots & \vdots & \vdots
 \end{array}$$

We can call this matrix ' Σ '. We can refer to each of the factual-moral worlds in Σ as ' σ '. This will parallel the standard practice of calling a set of possible worlds ' U ', and referring to each of the possible worlds in U as ' w '. The

crucial advantage to this kind of approach to the embedding problem is that it starts off by clearly specifying the character of the semantic objects which correspond to the making of FOMS. In the version of noncognitivism I am espousing, these semantic objects are the ordered pairs $\langle w, UMP \rangle$ depicted in the matrix. This set of factual-moral worlds plays the same role that a set of possible worlds plays in conventional semantics. It provides us with the basis for constructing a noncognitivist logic which neatly parallels standard classical logic. To illustrate this, we can begin by considering the nature of conventional semantics.

According to Peter Schotch:

Standard semantics works like this: each and every sentence A will be assigned a semantic value relative to each and every 'semantic structure' or model. The structures to which allusion has just been made consist of pairs of the form (U, v) where U is one of the sets trotted out above (now called the underlying set of the model) and v is a function (called the assignment function of the model) which assigns to every simple sentence a subset of U .²¹

Since the underlying set of my noncognitivist semantics consists of the factual-moral worlds contained in Σ , the semantic structure of noncognitivist semantics will consist of pairs of the form (Σ, v) . This structure can readily avail itself of results which are parallel to those already achieved by standard semantics. Standard semantics allows us to provide a rigorous definition of 'truth' by defining truth conditions in terms of possible worlds, the objects in its underlying

²¹Schotch, 1992, p.1

set. The assignment function of the model, v , serves as a truth function when it assigns to every simple sentence a subset of U . This assignment specifies the set of possible worlds within which that classical sentence is true. This can be formally characterized as follows:

$$v(A) \subseteq U; \text{ ie. } v(A) = \{w_a, w_b, \dots, w_z\}.$$

Together, all of these truth assignments provide an extensional definition of 'truth'. My noncognitivist semantics allows us to provide a rigorous definition of 'defensibility' by defining defensibility conditions in terms of factual-moral worlds, the objects in its underlying set. The assignment function of the model, v , acts as a defensibility function when it assigns to every simple FOMS a subset of Σ . This assignment specifies the set of factual-moral worlds within which that FOMS is defensible. This can be formally characterized as follows:

$$v(A) \subseteq \Sigma; \text{ ie. } v(A) = \{\sigma_a, \sigma_b, \dots, \sigma_z\}.$$

Together, all of these defensibility assignments provide an extensional definition of 'defensibility'.

In both systems, we can then describe a set of recursion clauses that will define the logical operators in terms of the assignment function. The set of recursion clauses required for sentential logic that accrue from standard semantics are as follows:

$$\text{Basis: } \models_v^w A \Leftrightarrow w \in v(A), \text{ if } A \text{ is atomic}$$

$$[T\wedge] \models_v^w A \wedge B \Leftrightarrow \models_v^w A \text{ AND } \models_v^w B$$

$$[TV] \models_v^w AVB \Leftrightarrow \models_v^w A \text{ OR } \models_v^w B$$

$$[T\rightarrow] \models_v^w A\rightarrow B \Leftrightarrow \models_v^w A \Rightarrow \models_v^w B$$

$$[T\sim] \models_v^w \sim A \Leftrightarrow \text{NOT } \models_v^w A.^{22}$$

The set of recursion clauses required for sentential logic that accrue from noncognitivist semantics are as follows:

Basis: $\models_v^{\sigma} A \Leftrightarrow \sigma \in v(A)$, if A is atomic

$$[D\wedge] \models_v^{\sigma} A\wedge B \Leftrightarrow \models_v^{\sigma} A \text{ AND } \models_v^{\sigma} B$$

$$[DV] \models_v^{\sigma} AVB \Leftrightarrow \models_v^{\sigma} A \text{ OR } \models_v^{\sigma} B$$

$$[D\rightarrow] \models_v^{\sigma} A\rightarrow B \Leftrightarrow \models_v^{\sigma} A \Rightarrow \models_v^{\sigma} B$$

$$[D\sim] \models_v^{\sigma} \sim A \Leftrightarrow \text{NOT } \models_v^{\sigma} A.$$

Both sets of recursion clauses allow us to apply the assignment function to sentences that are not simple - that is to say, sentences which contain logical operators. In doing so, these recursion clauses act as the semantic foundation which undergirds the operator rules of conventional sentential logic. Since the operator rules are the same, the set of inference patterns that are licensed by noncognitivist semantics will be exactly the same as the set of inference patterns that are licensed by standard semantics.

But there is an important difference between classical logic and noncognitivist logic which results from their differing semantics. This crucial difference is as follows. In standard semantics, we will take ' \models_v^w ' to mean 'true in possible world w with respect to assignment v'. That is to say, if $\models_v^w A$, then A is true in possible world w as specified

²²Loosely borrowed from Schotch, 1991, Ch.2, p.9.

by assignment function v . Thus, in standard semantics, the recursion clauses function as truth-conditions. In noncognitivist semantics, we will take ' \models_v^{σ} ' to mean '**defensible** in factual-moral world σ with respect to assignment v '. That is to say, if $\models_v^{\sigma} A$, then A is defensible in factual-moral world σ as specified by assignment function v . Thus, in noncognitivist semantics, the recursion clauses function as defensibility-conditions.

Having noted this difference in the character of the recursion clauses in classical logic and noncognitive logic, we can provide parallel characterizations of the key logical notion of **entailment**, which is symbolized by ' \models '. In standard semantics, we have:

$$\Gamma \models A \Leftrightarrow \forall w \in U, \forall v: (\models_v^w \Gamma \Rightarrow \models_v^w A).$$

This will read: 'The set of sentences Γ entails the sentence A , if and only if, whenever all the sentences in Γ are true, A is true as well'. In noncognitive semantics, we have:

$$\Gamma \models A \Leftrightarrow \forall \sigma \in \Sigma, \forall v: (\models_v^{\sigma} \Gamma \Rightarrow \models_v^{\sigma} A).$$

This will read: The set of sentences Γ entails the sentence A , if and only if, whenever all the sentences in Γ are defensible, A is defensible as well.

This is all the semantic machinery that is required to undergird the relation of **implication**. Implication is the notion we encounter in everyday arguments, such as the variations of modus ponens we have been discussing throughout this chapter, where a set of sentences is offered as grounds

for another sentence. In such cases, the claim is that the latter sentence 'follows from' the former sentences. To say that D follows from A, B, and C is to say that one is justified in inferring D from A, B, and C. In more technical parlance, it is to say that a rule which licences the assertion of D on the basis of A, B, and C is a correct rule of reasoning. The parallel versions of semantic machinery we have just described can be used to justify correct rules of inference in both standard and noncognitive cases. An important thing to notice is that the parallel semantic machinery allows us to give the same characterization of the intuitive meaning of implication or 'follows from' in both the standard and the noncognitive instances. This is just what we would expect.

We can return to Blackburn's modus ponens example to illustrate the parallel approaches. We will begin by considering the way that standard semantics could be used to justify this inference pattern. Presumably, this procedure would be favored by the moral realist. According to this approach, the example should be semantically construed as follows:

(1c) 'It is wrong to tell lies' is true in w ,
for $\forall w, \forall v: \models_v^w L$.

(2c) 'If it is wrong to tell lies, it is wrong to get your little brother to tell lies' is true in w ,
for $\forall w, \forall v: \models_v^w L \rightarrow B$.

(3c) 'It is wrong to get your little brother to tell lies' is true in w ,
 for $\forall w, \forall v: \models_v^w B$.

According to standard semantics, this inference is valid because it is **truth-preserving**. It is truth-preserving because it is not possible to have a w where (1c) and (2c) are true and (3c) is false. This is required by the third recursion clause of standard semantics:

$$[T \rightarrow] \models_v^w A \rightarrow B \Leftrightarrow \models_v^w A \Rightarrow \models_v^w B.$$

The preceding demonstration can then be generalized to provide semantic justification for all of the inferences licensed by the operator rules of traditional sentential logic.

Of course, this solution is not open to the noncognitivist because it requires us to treat moral sentences as admitting of truth or falsity. But a parallel procedure is open to her. According to noncognitivist logic, the example should be semantically construed as follows:

(1d) 'It is wrong to tell lies' is defensible in σ ,
 for $\forall \sigma, \forall v: \models_v^\sigma L$.

(2d) 'If it is wrong to tell lies, it is wrong to get your little brother to tell lies' is defensible in σ ,
 for $\forall \sigma, \forall v: \models_v^\sigma L \rightarrow B$.

(3d) 'It is wrong to get your little brother to tell lies' is defensible in σ ,
 for $\forall \sigma, \forall v: \models_v^\sigma B$.

According to noncognitivist semantics, this inference is valid because it is **defensibility-preserving**. It is defensibility-preserving because it is not possible to have a σ where (1d) and (2d) are defensible and (3d) is not. This is required by the corresponding third recursion clause of noncognitivist semantics:

$$[D\rightarrow] \models_{\forall}^{\sigma} A \rightarrow B \Leftrightarrow \models_{\forall}^{\sigma} A \Rightarrow \models_{\forall}^{\sigma} B.$$

As in the case of standard semantics, the preceding demonstration can be generalized to provide semantic justification for all of the inferences licensed by the operator rules of traditional sentential logic. The moral of the story is that it is no more difficult for the noncognitivist to justify inferences containing moral language than it is for the moral realist. A direct solution to the embedding problem is available.

3.9: Validity And Normativity: Two Separate Issues

At this point, some readers might have the following lingering question: What has happened to normative force? I am taking 'normative force' here to refer to the noncognitive compulsion which people generally experience to conform their behavioral patterns to the FOMSs they endorse. It appears that the preceding scheme can only account for the logical properties associated with sets of FOMS at the cost of removing their intrinsic normative force. If the noncognitivist semantics which undergirds inference patterns containing FOMS only uses 'defensible' in a purely descriptive

sense, what is the source of the FOMS's normative force? It can't be the mere fact that the FOMS happens to be defensible (in the descriptive sense) in a particular set of factual-moral worlds. No normative implications follow from this. Where does the FOMS's normative force come from? The answer is that the **normative force** of a FOMS comes from the UMPs within which it is descriptively defensible, given that some set of empirical conditions is taken to obtain.

A person takes the behavioral demands associated with certain FOMSS seriously because she embraces a particular UMP. In order to sincerely endorse a FOMS, she must embrace an UMP within which that FOMS is, as a matter of fact defensible. The fact that the FOMS is defensible within the UMP she embraces has important normative implications **for her**. For her, this fact provides the FOMS in question with normative force. By her lights, that FOMS's being descriptively defensible in the UMP she embraces makes that FOMS **defensible** in a normative sense. Thus, all normative force associated with the FOMS ultimately accrues from the UMP she embraces. All those who also embrace an UMP within which the FOMS is descriptively defensible will also take the FOMS to be normatively defensible. But this is a separate issue from the assessment of the logical properties of FOMSS. The important logical point is that even a person who does not take a particular set of FOMSS to have normative force because he does not embrace an UMP within which these FOMSS are descriptively defensible

can still see the logical relations that hold between them. It is not necessary to assent to a set of FOMSS to understand their implications. Consider the following argument:

Homosexuality is wrong.

If homosexuality is wrong, then same-sex marriages should be prohibited.

Therefore, same-sex marriages should be prohibited.

As a result of the UMPs they embrace, some people will take the conclusion of this argument to have normative force. Since they take both of the premises to be defensible in the normative sense, and they see that the pattern of inference is valid, they will take the conclusion to be defensible in the normative sense as well. By their lights, this is a **sound** argument. Other people will take the conclusion of this argument to have no normative force, as a result of the UMPs they embrace. Since they do not take the premises to be defensible in the normative sense, they will not take the conclusion to be defensible in the normative sense even though they can see that the pattern of inference is valid. By their lights, this is a valid argument that is **not** sound. The preceding discussion explains this in a way that abstracts from the normative force which anyone who sincerely endorses a set of FOMSS will naturally take those FOMSS to possess. But this is not a shortcoming of such an explanation. Explaining the normative force people attribute to FOMSS and explaining the logical properties of FOMSS are two separate issues. The

contemporary noncognitivist is able to do both of these. Thus, the embedding problem is not a reason to reject the version of noncognitivism I espouse.

CHAPTER FOUR: MORAL EXPLANATIONS

4.1: Introduction

Having provided a preliminary argument in favor of noncognitivism in the first chapter, I went on to explicate the nature of the particular version of noncognitivism I espouse more fully. From there, I defended this version of noncognitivism against the two principal criticisms levelled against traditional noncognitivism: the amoralist argument and the embedding problem. In this, the final chapter, I will abandon my defensive posture and end the dissertation by going on the attack. The focus of my campaign will be Nicholas Sturgeon's influential discussion of the existence of objective moral properties. We can call this 'the explanation defense'. This discussion represents one of the earliest defenses of a new conceptual basis for the doctrine of moral realism. As such, it has been very influential in the development of Cornell Moral Realism. We can begin the chapter by considering some of the philosophical background issues that are associated with the explanation defense.

One important consideration is the relative lack of respect which certain philosophers have afforded to the philosophical consideration of moral issues. Despite the fact that many contemporary philosophers are becoming increasingly open to alternative perspectives on philosophical issues, a peculiar kind of chauvinism is still a pervasive aspect of

academic philosophy. This is especially true in analytic circles. I think we're all familiar with what I'm talking about here. According to this exclusivist perspective, only certain members of the philosophical community are engaged in the practice of **real** philosophy. These are, of course, those who are engaged in such mainline pursuits as epistemology, philosophy of language, philosophy of mind, logic, and, especially, the philosophy of science. Those who lack the intellectual wherewithal to do **real** philosophy spend their time engaged in such dubious pursuits as ethics, or, even worse, aesthetics. Here, exactness and cogent argumentation are passed over in favor of clever rhetorical stratagems and empty appeals to emotion. The problem is that since there is no clear cut subject matter we can zero in on, there is no definitive way of determining whose opinion is correct and whose isn't in these kinds of discussions. Objectivity is cast by the wayside. The end result is a bunch of people with expensive educations espousing their opinions about the way they think the world should be. At the end of the day, the person with the cleverest theory wins. All we really have here is an exercise in intellectual masturbation. It might make the participants feel good, but it's not going to yield any fruitful results. As serious as such discussions may seem to be, they can hardly be counted as **real** philosophy.

The explanation defense offers an interesting reply to the preceding perspective. It accepts the terms of discussion

presupposed by the chauvinistic picture of what **real** philosophy must be. But it rejects the claim that ethics does not fall under the purview of **real** philosophy. It attempts to legitimize ethical theorizing by arguing that the way we gain ethical knowledge is strongly analogous to the way that we gain scientific knowledge. It holds that ethical claims can be substantiated with the same dispassionate rigor as scientific claims. Just as there are scientific facts which are there to be discovered by those who employ the appropriate means to detect such facts, so there are ethical facts which are there to be discovered by those who employ the appropriate means to detect them. According to the explanation defense, there is a distinct subject matter to which ethical theorizing pertains. When people are engaging in a moral dispute, they are arguing about the way that the world is just as they are when they are engaged in a scientific dispute. As in the case of a scientific dispute, the way that the world turns out to be will determine which of the two disputants is actually correct. There is something in the world which will determine whether the moral opinions people express are true or not. Moral experts are those people who are best able to ascertain the nature of these truth conditions. If we understand the situation correctly, we will see that ethical discourse is grounded in reality; it is more than the simple flaunting of groundless opinions. If the explanation defense is successful, then ethics will have been placed on the same solid ground as

the philosophy of science. As such, it can finally take its rightful place as a respectable part of **real** philosophy.

In the course of this chapter, I shall attempt to show that this grand attempt to put ethical theorizing on a firmer footing is ultimately doomed to failure. The reason it is doomed to failure is that it implicitly assumes that the full sense of what is expressed by our ethical discourse can be adequately characterized in terms of truth conditions. In the first chapter, we saw that this assumption is not plausible. Since the full sense of what is expressed by our ethical discourse cannot be adequately characterized in terms of truth conditions, there is no set of truth conditions which can serve as the standard for determining which moral claims are acceptable and which are not. Since this is so, the possibility of looking to a set of natural properties to provide the truth conditions for moral claims disappears. Thus, it is a mistake to think that ethical inquiry can be modeled after scientific inquiry. Those who think it can fail to appreciate the character of moral language use. Or so I shall argue. The question of whether or not ethical theory is a part of **real** philosophy, or why this should matter, I leave to others.

A second philosophical background issue associated with the explanation defense is an important development in recent philosophy that has come to be known as 'naturalized epistemology'. The key feature of this new movement is a

change in the focus of epistemological assessment. While traditional epistemology concentrates on **beliefs**, naturalized epistemology focuses on **believers**. Traditional epistemology is concerned with providing a formulation of the grounds which will provide any believer with an appropriate justification for assenting to a particular proposition or set of propositions. Naturalized epistemology is concerned with providing a description of the conditions and mechanisms which will result in a reliable process of belief formation for human believers. As such, it is intimately linked with causal theories of knowledge. Hilary Kornblith describes such theories as follows:

Causal theorists of knowledge believe that knowledge is reliably produced true belief. ... Causal theorists of knowledge do not deny that knowledge is some sort of justified true belief; they merely give a nonstandard account of what it is for a belief to be justified. They claim that a belief is justified just in case it is caused by a reliable process. ... In claiming that knowledge is reliably produced true belief, causal theorists of knowledge are thus not abandoning the notion of justification, but rather committing themselves to a reliabilist account of justification.¹

4.2: Epistemology Naturalized In Morality And Science

Gilbert Harman was the first philosopher to apply this new perspective on epistemological issues to moral theory. According to Harman 'the problem with ethics' is that moral facts have no role to play in reliabilist accounts of human knowledge acquisition. Nicholas Sturgeon agrees with Harman's general approach to issues of knowledge here, but disagrees

¹Kornblith, 1985, p.116.

with him on precisely this point. Sturgeon claims that moral facts do have a role to play in reliabilist accounts of knowledge acquisition. Thus, the locus of their disagreement centers on one key issue: whether or not the existence of moral facts plays an essential role in the description of processes of reliable moral belief formation in human beings. It is important to notice that a defense of moral facts does not require a theory of the **direct** perception of moral facts such as we find in the early moral sense theorists. It only requires that moral facts are included in the complete ontological picture which provides the best description of reliable belief formation. It is not necessary for the defender of moral facts to provide an explanation of how moral facts are able to directly impinge on human sense organs in order to vindicate the existence of moral facts.

The debate between Harman and Sturgeon focuses on the comparison of scientific and moral scenarios. They disagree over the extent to which the two types of cases are parallel. Harman believes that there is an important difference between these two types of cases while Sturgeon does not. To illustrate this disagreement we can begin by considering a scientific scenario. Suppose that a scientist correctly believes that a proton has passed through a cloud chamber on the basis of his seeing a vapor trail. He directly perceives the vapor trail, but does not directly perceive the proton. Rather, he infers the existence of the proton as part of the

best explanation for the existence of the vapor trail which he does perceive. Belief in the existence of a proton is vindicated because the best causal story of how the scientist came to believe that a proton was present includes the presence of a proton. It is not possible to construct a plausible causal history of the scientist's belief which does not include the existence of a proton. This approach to the scientific case constitutes a specific application of a familiar epistemological strategy known as 'inference to the best explanation'. According to this strategy,

one infers, from the premise that a given hypothesis would provide a 'better' explanation for the evidence than would any other hypothesis, to the conclusion that the given hypothesis is true.²

Since the hypothesis that a proton passed through the cloud chamber provides the best explanation for the scientist's believing that a proton passed through the cloud chamber, it is thereby vindicated.

Both parties in this debate agree that the preceding scenario describes an instance of valid scientific reasoning. The old positivist view that the existence of protons can never be demonstrated because they are in principle unobservable is a nonstarter here since in principle observability is no longer regarded as a plausible prerequisite for admitting an entity into one's ontology. The bone of contention between Harman and Sturgeon centers on

²Harman, 1965, p.89.

whether or not this same basic approach can be applied to the moral realm in order to vindicate the existence of moral facts. According to Sturgeon, as in the case of protons, the existence of moral facts will be vindicated since it is necessary to posit their existence in order to provide the best explanation for certain common phenomena.

Harman argues that the existence of moral facts cannot be established in this way. He begins his case by considering a moral scenario where, as a passerby, "you round a corner and see a group of young hoodlums pour gasoline on a cat and ignite it."³ You will naturally deem this act to be morally reprehensible. How are we to explain this reaction of yours? Harman claims that your reaction can be explained entirely in terms of psychological facts about you. The best explanation of your being morally repulsed at the sight of the young men setting fire to the cat is that you are psychologically disposed to be repulsed at displays of senseless cruelty. It is not necessary to posit, in addition, the moral fact that cat-burning is wrong in order to explain the occurrence of your experience of moral revulsion. As Harman claims:

Indeed, an assumption about moral facts would seem to be totally irrelevant to the explanation of your making the judgment you make. It would seem that all we need assume is that you have certain more or less well articulated moral principles that are reflected in the judgments you make, based on your moral sensibility. It seems to be

³Harman, 1977, p.4.

completely irrelevant to our explanation whether your intuitive immediate judgment is true or false.⁴

If Harman is correct, then we cannot use our experience of making spontaneous moral judgments as a possible basis for inferring the existence of moral facts. As a result of our psychological constitutions, we would have the same reactions to actions like cat-burnings regardless of whether or not it is a fact that such actions are wrong.

4.3: Moral Explanations

Nicholas Sturgeon presents his well-known criticism of Harman's position in an article entitled 'Moral Explanations'. Recall that, in essence, Harman claims that the problem with ethics is that moral principles cannot be tested against the world the way that scientific theories can. For Sturgeon, scientific theories and moral principles are much more on a par than people think. He thinks that moral principles can be tested in the same way as scientific principles. Sturgeon's main line of attack is to take Harman to task on the claim that putative moral facts would be 'explanatorily irrelevant' in arriving at an adequate explanation of our moral experiences.

It should be emphasized that the issue at this point in the debate is not whether or not there are any moral facts. The issue is whether moral facts, if there were any, would play a role in a reliabilist account of moral knowledge.

⁴Harman, 1977, p.7.

Harman claims that they would not. His thesis is that a distinctive problem with the notion of moral facts is that even if there were any, they would play no role in any reliable processes of human detection. Sturgeon denies this. In order to challenge Harman's claim, it is entirely appropriate for Sturgeon to provisionally assume that there are moral facts. As he notes:

it should be clear that the assessment of this [Harman's] thesis not merely permits, but **requires** that we provisionally assume the existence of moral facts. I can see no way of evaluating the claim that **even if** we assumed the existence of moral facts they would still appear explanatorily irrelevant, without assuming the existence of some, to see how they would look.⁵

In characterizing Harman's position, Sturgeon distinguishes between two types of moral phenomena which require explanation. He says:

So the claim is that moral facts are not needed to explain our having any of the moral beliefs we do, whether or not those beliefs are observations, and are equally unneeded to explain any of the observations we make, whether or not those observations are moral.⁶

Sturgeon goes on to claim that moral facts are needed to explain each of these two aspects of our experience. On the one hand, the fact that the burning of the cat is wrong is necessary to explain our believing that it is wrong. On the other hand, the fact that the boys possess certain moral character qualities is necessary to explain our observing them treat the cat in the way that they do. We might say that there

⁵Sturgeon, 1984, p.57.

⁶Sturgeon, 1984, p.54.

are two varieties of moral facts involved here. The first kind are those moral facts which cause people to have certain kinds of observational experiences. The second kind are those moral facts which cause people to do certain kinds of deeds. For convenience, I will refer to the first type as observation-causing moral facts and the second type as deed-causing moral facts.

Focusing on Harman's claim that "assumptions about moral facts seem 'completely irrelevant' in explaining moral observations and moral beliefs", Sturgeon says that:

it is natural to think that if a particular assumption is **completely irrelevant** to the explanation of a certain fact, then the fact would have obtained, and we could have explained it just as well, even if the assumption had been false.⁷

What Sturgeon is proposing here is a **counterfactual test** for explanatory relevance. According to this test, we can determine whether the presence of **x** is relevant to the explanation of **y** by considering whether **y** would still have occurred even if, counter to fact, **x** had not been the case. If **y** still occurs in all of the closest possible worlds where **x** is not the case, then **x** is explanatorily irrelevant for **y**. If **y** does not occur in all of the closest possible worlds where **x** is not the case, then **x** is explanatorily relevant for **y**. Given any explanandum **y**, we are only justified in inferring the existence of those explanans **x** which are explanatorily relevant for **y**'s being the case. Sturgeon claims that when we

⁷Sturgeon, 1984, p.65.

apply the counterfactual test to moral properties we will find that they are indeed explanatorily relevant.

4.4: Explanatory Relevance And Explanatory Potence

As Sayre-McCord points out, however, the inferring power of Sturgeon's counterfactual test is too strong. Its general employment commits us to the existence of certain properties which it does not seem plausible to endorse. Suppose someone claims that witches exist and that the property of being a witch supervenes on some set of natural properties which are possessed by all and only people who are witches. This set of natural properties might include having an average specific gravity of less than one. She could then claim that part of the explanation of Hilda's floating when she was bound and thrown into a pond is that Hilda is a witch. She would go on to point out that if Hilda were not a witch, she would have to have an average specific gravity of greater than one. But, if this were the case, Hilda would not have floated when she was bound up and thrown into the pond. According to Sturgeon's counterfactual test, the claim that Hilda is a witch has been vindicated as part of the best explanation of her floating when thrown in the pond. In order to counteract this kind of move, Sayre-McCord proposes that we adopt a criterion of **explanatoryotence**, rather than mere **explanatory relevance**, to act as a constraint on which entities can be legitimately posited by an inference to the best explanation strategy. He describes this criterion as follows:

a particular assumption is explanatorily impotent with respect to a certain fact if the fact would have obtained and we could have explained it just as well **even if the assumption had not been invoked in the explanation** (as opposed to: "even if the assumption had been false").⁸

While Sturgeon's counterfactual test only rejects the inference of entities which are explanatorily irrelevant, Sayre-McCord's counterfactual test also rejects the inference of entities which are explanatorily impotent. As such, it allows us to reject the preceding inference that Hilda is a witch on the grounds that such an inference is explanatorily impotent. Hilda's sinking in the pond can be explained entirely in terms of the specific gravity of her body mass without any mention of whether or not she is a witch. Pointing out this additional factor adds nothing to the explanation. Thus, its inference is illegitimate if we adopt a criteria of explanatory potency.

Though Sayre-McCord's point is well-taken, it does not diminish the thrust of Sturgeon's main point vis-à-vis Harman. This is that moral facts are an essential part of the explanatory ontology which best accounts for our moral experience. This is the basis upon which Sturgeon advocates the positing of moral facts. Harman is making the strong claim that even if there were moral facts, they would be explanatorily irrelevant. Sturgeon is arguing against this strong claim. He is attempting to show that moral facts are explanatorily **relevant**. If he can do this, then he will have

⁸Sayre-McCord, 1988, p.272.

refuted Harman's case for rejecting moral realism. Whether moral facts are part of the best explanatory ontology because of their explanatory relevance or because of their explanatory potency is not an issue in Sturgeon's debate with Harman. This is an issue of theoretical refinement which is most appropriately discussed among those who are firmly ensconced in the moral realist camp. Having taken note of Sayre-McCord's argument in favor of adjusting Sturgeon's position, we can return the debate between Sturgeon and Harman.

4.5: Supervenience And Necessity

Consider the cat-burning scenario again. According to the counterfactual test, the observation-causing moral fact that the hoodlums' actions are wrong is only explanatorily irrelevant to our being repulsed by them if we would have still been repulsed by them even though the hoodlums' actions were not wrong. So, in order to determine whether or not the observation-causing moral fact that the hoodlums' actions are wrong is explanatorily irrelevant to our being repulsed by them, we need to consider scenarios where their actions are not wrong to see what our reactions would be. It is crucial to notice that there are certain constraints on how this set of closest possible worlds can be evaluated. If moral properties supervene on or are constituted by natural properties, then there is no possible world where their actions are not wrong which does not also differ in some of its natural properties

from a world in which their actions are wrong. Since this is the case:

in order to imagine them not doing something wrong we are going to have to suppose their action different from the actual one in some of its natural features as well. So our question becomes: Even if the children had been doing something else, something just different enough not to be wrong, would you have taken them even so to be doing something wrong?⁹

The claim here is that certain scenarios are simply not intelligible. This is based on the following general principle of supervenience relations which is by no means limited to the moral case.

If the base properties in two contexts are all physically identical, then the supervening properties must be identical as well.

The contrapositive of this general principle tells us that if the supervening properties in two contexts are not identical, then their base properties cannot all be identical. We can illustrate this principle with a simple scientific case. Suppose you are given the task of scientifically examining a small lump of gold. Later, you are allowed to hold what appears to be the same lump. It turns out, however, that the first lump was replaced with another lump which is identical in size, shape, color, and mass. Upon closer inspection, you discover that this second lump also has the same atomic weight, specific gravity, malleability, and conductivity. The reason for this is that the second lump has been manufactured

⁹Sturgeon, 1984, p.66.

by a special replicating machine, which made it physically identical to the first lump right down to its atomic microstructure. The only difference is that the first lump is genuine gold but the second lump isn't. This scenario is not intelligible. If the first lump was gold and the second lump has an identical microstructure, then the second lump has to be gold as well. The only way for the second lump not to be gold is for it to have a different microstructure from the first lump. In the same way, the only way for the hoodlums' actions not to be wrong in alternative world scenarios would be for them to differ in some of their natural base features. It is simply not intelligible to imagine a situation where the hoodlums' actions are the same in all their natural features, yet are not wrong. The supervenient, observation-causing moral fact that the hoodlum's actions are wrong can only obtain within a specific range of possible worlds. These are the only worlds that count when we are evaluating whether or not this fact is explanatorily relevant to our thinking that their actions are wrong.

4.6: Six Aspects Of Comparison

The point Sturgeon is getting at can be made clearer if we consider the elements of similarity between the scientific and the moral cases more closely. Recall that he is attempting to convince us that an inference to the best explanation strategy licences us to posit the existence of moral properties for the same reason that it licences us to posit

the existence of unobservable physical entities. It seems to me that there are six different aspects of the comparison between the moral and the scientific case which need to be considered here.¹⁰ First, there is the psychological profile of the observer. In both the scientific and the moral case, the observer is predisposed to make judgments in accordance with a particular theory. Second, there is the issue of the veracity of the theory which the judge is predisposed to rely on. Third, there is the catalytic event which results in the spontaneous judgment. Fourth, there is the *ceteris paribus* assumption that the observation is being made under 'normal' circumstances. Fifth, there is the obtaining of the fact which is reported by the observer. Finally, there is the explanandum; the observer makes a report. The first five elements **together** constitute the best explanation for the observer's making the report she did. To be sure, these five elements only explain why the observer is in an intentional state which corresponds with the report. If, for simplicity's sake, we assume that the observer is also predisposed to make sincere reports, these five elements will explain why the report was made.

It is crucial to notice that the fifth element has a legitimate place in the explanatory picture only so long as the other elements hold. If any of the first four elements of

¹⁰ This characterization is loosely based on a four point comparison mentioned by Sayre-McCord in Sayre-McCord, 1988, p.270.

the explanans fail, the fifth element is no longer required to explain the observer's report. Thus, inference to the existence of the fifth element is only warranted within a limited range of contexts where a number of presuppositions already hold. In both the moral and the scientific scenarios, the claim is: given that certain facts are already stipulated to be part of the explanatory picture in a particular case of the formation of an intentional state which leads to the making of a report, an additional fact must also be part of that same explanatory ontology. If the set of initial stipulations is altered in any way, then, the necessity of the fifth element immediately falls into question.

To illustrate this, I will begin by considering the explanatory role played by the first two elements of the explanans in the scientific case. They are best considered together since their roles are interconnected. In this case, the scientist is predisposed to report that a proton was present whenever he sees a vapor trail in a cloud chamber. This is his psychological profile. It is a result of his scientific training and distinguishes him from the layperson who is not predisposed to make such a report in similar circumstances. It leads him to make theory-laden observations. This theory-ladenness does not present a problem, however, as long as the scientific theory which predisposes the scientist to interpret events in such a manner is essentially correct. This is where the second element comes in. Since the

scientific theory which underlies subatomic experimentation through the use of cloud chambers is correct, the best explanation for the scientist's theoretically-motivated report that a proton was present must include the fact that a proton **was** present. This is an extra-textual fact about our world which the scientifically informed reader incorporates into the example. The theory-ladenness of the scientist's observation would only present a problem if the scientific theory underlying his disposition to make certain judgments was not correct. Suppose, for example, that we live in a world very much like the present one except that the theory underlying the use of cloud chambers in subatomic experimentation is not entirely correct. As a result of this theoretical inadequacy, approximately one percent of the vapor trails observed in the cloud chambers constructed here are actually caused by stray electrons. Since no one is looking for the presence of stray electrons in cloud chambers, however, these deviations tend to go completely unnoticed. The scientific community is under the mistaken impression that our present cloud chamber technology is completely reliable. If this were the case, then the presence of a vapor trail would not require that a proton had passed through the cloud chamber. Here, the theory-ladenness of the scientist's observation does present a problem. Since the scientist's use of the cloud chamber is an imperfect proton-detecting process, no particular instance of his reporting the presence of a proton in the chamber necessitates

the presence of a proton in the chamber. Any particular report could have been the result of an electron passing through the cloud chamber. Thus, a legitimate inference to the fifth element of the explanans is contingent on the first two elements holding.

Next, I will consider the explanatory role of the third and fourth elements of the explanans in the scientific case. These are also somewhat interrelated. The third element requires that the appropriate sort of catalytic event precedes the spontaneous report made by the observer. In the proton case, there must actually be a vapor trail in the cloud chamber. Suppose that after many long hours of staring at the screen and tabulating the occurrence of vapor trails, the scientist begins to experience a mild sort of hallucinatory effect. Every so often she 'sees' a vapor trail that is not really there and on this basis reports that a proton has passed through the chamber. Clearly, in these types of cases, a proton actually passing through the chamber is not part of the best explanation of her report. We can also imagine various circumstances involving lighting problems, marks on the screen of the cloud chamber, electrical glitches, and various other technical malfunctions with the equipment. Any of these would constitute a failure of the fourth element. It would produce the appearance of a vapor trail which was not in fact caused by a passing proton. Again, a proton passing through the cloud chamber is not part of the best explanation

of the scientist's report in this range of cases. We can see, then, that a number of factors must already be in place before the passing of a proton through the cloud chamber becomes a necessary constituent of the best explanatory ontology that accounts for the scientist's proton report.

We can now move on to consider the moral case. We need to keep in mind that in order to produce an example which shows that the moral case is relevantly disanalogous to the scientific one, one must describe a situation that meets two requirements. First, the situation must be similar to the scientific case in that the first four elements of the explanans are satisfied. Second, the situation must be different from the scientific case in that the fifth element of the explanans is not required. Examples of moral inferences which don't satisfy the first four elements and don't require the fifth element are unacceptable. They don't show that the moral case is distinct from the scientific cases previously discussed where the failure of one or more of the first four elements removes the need for the fifth element.

One inappropriate class of counterexamples alters the nature of the catalytic event which results in the spontaneous report. For example, Evan Jobe adduces a variant of Harman's cat-burning example where

some children are setting fire to a cushion in the form of a cat, perhaps for some good reason".¹¹

¹¹Jobe, 1990, p.270.

He claims that since the pattern of retinal stimulation in this case would be qualitatively identical to the pattern of retinal stimulation the observer would experience if the children really were burning a live cat, any fact about the wrongness of the children's actions can't function explanatorily. This does not follow, however. In Jobe's scenario, we don't need to posit the fifth element to explain the observer's report because the third and fourth elements have not been satisfied. We have a situation where the appearance of a burning cat was not caused by an actual burning cat. But the fact that we don't need to posit any moral fact about the boys' actions in this case does nothing to show that the positing of a moral fact is not necessary in the cases where they really are setting fire to a cat. And the latter range of cases are the only ones we are interested in considering. Sturgeon's point is that if we consider the limited range of moral scenarios which are exactly parallel to the limited range of scientific scenarios which satisfy the requirements for legitimately positing the existence of unobservable entity facts, we will see that we are justified in positing the existence of moral facts for this limited range of moral scenarios. All other cases are irrelevant.

We can now consider another kind of scenario which might be used in an attempt to show that the moral case is relevantly disanalogous to the scientific one. Imagine a middle-aged businessman from Mobile, Alabama who reacts with

moral indignation when he sees a young black man and a young white woman walking down the street holding hands. Isn't his phenomenological experience qualitatively identical to the moral revulsion experienced by the witness to the cat-burning? Clearly, it is not necessary to posit any moral fact about the nature of the couple's union to explain his reporting that what they are doing is wrong. All we need to point to are his segregationist dispositions. By the same token, one might claim that it is not necessary to posit a moral fact to explain the report made by the witness to the cat-burning either. Why do we need a moral fact in the latter case but not in the former?

According to the scheme we have been considering, the moral realist could attempt to defuse this counterexample by claiming that the second element fails here. That is, she could characterize the businessman's reaction as the result of his endorsing a flawed moral theory. This type of counterexample fails because it does not show that moral facts are not part of the best explanatory picture for moral judgments that are made from a true moral theory. As in the scientific case, the moral realist could claim that we are only interested in that range of cases where the observer's propensity to judge is based on an essentially correct theory. We want to know if facts corresponding to the observer's report are called for within **this** range. We need not concern

ourselves with reports made on the basis of flawed moral theories.

4.7: True Moral Theories

In order to be in a position to employ this defense, however, the moral realist must be able to articulate the nature of a true moral theory in a way that will provide a principled basis for distinguishing true moral theories from false ones. It seems to me that the explanation defense encounters a serious problem here. To understand the nature of this problem, we will need to retrace our steps in a way that emphasizes the subtle metaethical presuppositions which underlie the debate between Harman and Sturgeon. In general, this debate takes place within a cognitivist framework. The debaters presuppose that the moral realm is appropriately conceived of in terms of facts when they argue about whether or not moral facts have a legitimate place within an inference to the best explanation perspective. Harman endorses the sceptical conclusion that they do not while Sturgeon endorses the moral realist conclusion that they do. Sturgeon claims that only an ontology which includes moral facts can adequately account for our moral experiences. This ontology is contrasted with one which is entirely devoid of moral facts. He claims that if we seriously consider the nature of an ontology which contains no moral facts, we will see that it does not square with common moral phenomena which are in need of explanation. Given the assumption that our moral beliefs

are best construed as cognitive ones, Sturgeon presents an interesting case for the existence of moral facts. His claim is that moral theories are similar to scientific theories in important ways. According to Sturgeon, the best way to bring about an advancement in metaethical theory is to make it more scientific. Moral realism can be established by employing the methods of post-positivist realist philosophy of science to the moral realm. Though the scientific anti-realist will be unimpressed by such a ploy, those who find this commonsensical realist approach to the philosophy of science convincing will be apt to take Sturgeon's claim that this approach can be used to establish the existence of moral facts much more seriously.

In keeping with the cognitivist presuppositions that underlie his debate with Harman, an important part of Sturgeon's strategy for convincing the reader that moral theories are of a kind with scientific theories involves the **assumption** that there are **true** moral theories. He invokes our **normative** sensibilities in order to arrive at examples of true moral theories. Here are two extremely simple examples of true moral theories: (1) 'Acts of deliberate cruelty are wrong' and (2) 'People who wish to exterminate other people solely on the basis of their race are morally reprehensible'. On the basis of these putatively **true** moral theories, he claims that the hoodlums' burning of the cat was morally wrong and that Hitler was morally depraved. It is important be aware of what Sturgeon takes himself to be doing here. He is attempting to

provide a demonstration of the way that moral discourse parallels scientific discourse. In each of these two cases, he is claiming that a particular moral property is instantiated in a particular set of natural circumstances on the basis of a moral theory. The moral theory specifies the circumstances under which the moral property supervenes on the cluster of natural properties.

To illustrate the nature of Sturgeon's position, we can consider the explanatory role that each of the moral properties plays in its respective scenario. In the first scenario, the observation-causing moral fact that the hoodlums' action was wrong is a crucial part of the ontology which best explains your coming to make the judgment that their action was wrong. In the second scenario, the deed-causing moral fact that Hitler was morally depraved is a crucial part of the ontology which best explains why he did the things that he did. In order to demonstrate that each of the moral facts plays a crucial explanatory role in its respective scenario, Sturgeon invites us to consider alternative scenarios which are as close as possible to the original scenario except that the pertinent moral fact no longer obtains. A crucial difference between Sturgeon and Harman emerges in the way that such alternative scenarios are evaluated. As Sturgeon points out:

my approach differs from Harman's in one crucial way. For among the beliefs in which I have enough confidence to

rely on in evaluating explanations, at least at the outset, are some moral beliefs.¹²

The key difference between Sturgeon and Harman is that Sturgeon is willing to bring **moral background theories** into the initial assessment of moral scenarios. Since a moral theory specifies the circumstances under which moral properties supervene on natural properties, it will determine the entire set of instantiations of any particular moral property within any particular set of natural circumstances. The truth of the moral theory will make it **necessary** that the supervening moral property is present whenever the requisite natural base properties are present. As a result, certain statements about whether or not certain moral properties accompany certain sets of natural circumstances are **necessarily false**. For example, consider the antecedents of the following conditionals which correspond to the two scenarios under consideration:

[1] If the action [of burning the cat] were one of deliberate, pointless cruelty, but this **did not make it wrong**, you would still have thought it was wrong. ...

[2] If Hitler's psychology, and anything else about his situation that could strike us as morally relevant had been exactly as it in fact was, but **this had not constituted moral depravity**, he would still have done exactly what he did.

Sturgeon believes that the antecedents of these conditionals are both **necessarily false**.¹³

¹²Sturgeon, 1984, p.68.

¹³Sturgeon, 1984, p.69.

It seems to me that he is mistaken about this. Having brought Sturgeon's metaethical presumptions into the open, I shall now attempt to explicate what I take to be the nature of his mistake. At this point, it is important to recall that Sturgeon's entire discussion is premised on the assumption that there are moral facts. At a deeper level, this assumption is premised on the claim that the moral realm is comprised of facts. He has presupposed this in order to refute the claim that even if there were moral facts, they would play no part in the explanation of our moral beliefs and observations. He starts off by providing two extremely plausible candidates as examples of moral facts. If there are any moral facts at all, it seems like a good bet that these would be among them. The point I wish to emphasize is that one can be in complete **normative** agreement with Sturgeon concerning these two examples without sharing his **metaethical** position. One can also agree that Sturgeon has successfully refuted Harman's position without being forced to admit that his argument cogently establishes the existence of moral facts. All his argument establishes is that we will be committed to certain conclusions if we choose to speak about moral issues in a certain manner. If it is appropriate to speak in terms of moral **facts**, then Sturgeon's position is preferable to Harman's. If it turns out that our moral observations and experiences are not the sorts of things which are appropriately described as factual, however, then Sturgeon's

arguments will have done nothing to positively establish the existence of moral facts.

It is clear from Sturgeon's arguments that he does believe that it is appropriate to consider the moral realm as a realm of facts. Consider the following comments he makes about Hitler:

isn't it plausible that Hitler's morality - the fact of his really having been morally depraved - forms part of a reasonable explanation of why we believe he was depraved?¹⁴

This way of wording the question puts the gainsayer in a rather difficult position. Suppose I want to claim that the fact of Hitler's really having been morally depraved can't form part of a reasonable explanation of why we believe he was depraved because it simply isn't a fact that Hitler really was morally depraved. The likely response to this claim is that either I am seriously normatively misguided or I am saying something that I really don't believe just to argue for the sake of argument. Such a response would interpret the claim that it isn't a fact that Hitler really was morally depraved as tantamount to the claim that it is a fact that Hitler wasn't morally depraved. After all, either he was morally depraved or he wasn't. If I deny that he was, I am thereby committed to the claim that he wasn't. In like manner, it is either a fact that the cat is on the mat or it is a fact that it isn't. If I deny that it is a fact that the cat is on the

¹⁴Sturgeon, 1984, p.54.

mat, I am thereby committed to endorsing the claim that it is a fact that the cat isn't on the mat. If, indeed, these were the only two options available in the Hitler example, I, personally, would opt for the choice that it is a fact that Hitler was morally depraved. But these are not the only two options. I can deny that it is a fact that Hitler was morally depraved on the metaethical grounds that such a statement is an evaluative judgment about Hitler rather than a factual one. Sturgeon's response to Harman is framed in such a way that it cuts off this third possibility. It presupposes that a noncognitive response to this issue is not an option. As a result, **normative** conclusions that many of us find irresistible subtly end up committing us to a substantial **ontological** position. According to this way of conceiving the issue, for any normative conclusion one endorses, she must be willing to introduce the corresponding moral fact into her ontology. Conversely, denying that a particular moral fact exists requires the denial of the corresponding normative claim. According to this point of view, the entire array of moral facts which exist in the world can be ascertained simply by arriving at the correct moral theory.

The problem with framing the terms of the debate in this way is that the cards have already been stacked in favor of the moral realist. Unless one is willing to endorse an unpopular normative position, she is forced to accept the existence of moral facts. As such, the debate has conveniently

sidestepped the real issue here. The real issue is whether or not it is appropriate to speak of the moral realm as one which is constituted of matters of fact. If it is not, then moral statements are not the sorts of things which can be true in any correspondence sense since there are no corresponding states of affairs which can make them true or false. If this is so, then it is no longer obviously appropriate to speak of moral theories as being true either. In my estimation, the central problem with the debate between Harman and Sturgeon lies in the terms in which the whole discussion has been tacitly couched. No critical consideration has been given to the issue of whether or not it is appropriate to construe the moral realm as a factual one. Harman assumes that there is nothing initially problematic about the notion of moral facts. The problem is just that there are no legitimate grounds for positing their existence. Sturgeon counters by arguing that there are grounds for positing their existence. Like Harman, he assumes that there is nothing inherently problematic with the notion of moral facts in the first place. I, however, claim that there is something inherently problematic about referring to the moral realm as a factual one. Sturgeon's essay misses this problem because it focuses too quickly on the normative issue of what the moral facts might be without considering the ontological issue of what would have to be the case in order for something to be a moral fact. Thus, it does

not provide the irrefutable case in favor of the existence of moral facts that many advocates of moral realism suppose.

4.8: The Second And Fifth Explanans Reconsidered

In order to illustrate the difficulty Sturgeon faces here, I will return to the six point comparison between the scientific and the moral cases discussed earlier, focusing on the relationship between the second and fifth elements of the explanans. To facilitate this, I will **assume** that all of the remaining elements of the explanans hold for the remainder of this discussion. Recall that the second element concerns the truth of the theory the judger is predisposed to rely on, while the fifth refers to the obtaining of the fact which is reported by the observer. **If we suppose that all of the other elements of the explanans hold**, an asymmetrical dependency between the second and fifth elements emerges. The asymmetry is as follows. If the fact reported by the observer does not obtain, then the theory she relies on cannot be true. On the other hand, if the theory she relies on is not true, it is still possible for the fact she reports to obtain. Contrapositively, if the theory she relies on is true, the fact she reports must obtain. Yet, the obtaining of the fact she reports does not necessitate the truth of the theory she relies on. The reason for this is the familiar one that, *ceteris paribus*, accurate theories cannot make false predictions but erroneous theories can still make true predictions. A true prediction does not conclusively verify a

theory, but a false prediction conclusively falsifies a theory. The preceding comments are predicated on the assumption that the other elements of the explanans hold, of course.

Having described the asymmetrical relationship between true theories and factual reports that holds in certain contexts, we can consider two questions relating to true theories:

- (1) What is it that makes a true theory true?
- (2) What is it about a true theory in virtue of which the facts it posits or predictions it makes **must** obtain?

I will begin by considering scientific theories. We can use the proton case as an example. In the proton case, there is an underlying theory about how the ultimate constituents of the universe interact with each other which forms the basis for cloud chamber technology. This theory is true because the ultimate constituents of the universe interact in the way specified by the theory. They could have interacted differently, but, as a matter of fact, they don't. Since the theory that cloud chamber technology is based upon describes the world accurately, it is **physically impossible** for there to be a vapor trail in the cloud chamber which was not caused by a proton passing through the chamber. The mechanism is designed in such a way that it only allows vapor trails to be created by stray protons. The reason the theory cannot make false predictions, then, is that the underlying constituents

of the universe **really do** interact in certain ways and not in others. This makes it physically impossible for the ultimate constituents of the universe to interact in ways which would result in a false prediction for the theory that underlies cloud chamber technology. The universe is constituted in such a way that the predictions specified by cloud chamber technology **will** obtain. Cloud chamber technology is reliable **because** the universe is constituted in this way. Thus, in the scientific case, we can see that the two questions about true theories are interrelated. The same thing that makes a scientific theory true also makes it physically impossible for that theory to make false predictions. This 'thing' is simply the way that the world is.

From here, we can go on to consider these two questions as they relate to the moral realm. With regards to the first question, the determination of what it is that makes true moral theories true is not nearly such a straightforward matter in the moral case as it is in the scientific case. What does it actually **mean** to say that a moral theory is true? Surprisingly, Sturgeon has nothing to say about this important theoretical issue. Peter Railton, another prominent moral realist, provides a semblance of an answer to this question that is compatible with a naturalistic ontology. Naturalism here is the view that anything that exists is ultimately composed of physical components. Railton asks us to let him:

introduce an idealization of the notion of social rationality by considering what would be rationally

approved of were the interests of all potentially affected individuals counted equally under circumstances of full and vivid information. Because of the assumption of full and vivid information, the interests in question will be objective interests. ... [T]his idealization is equivalent to what is rational from a social point of view with regard to the realization of intrinsic non-moral goodness. This seems to me to be a recognizable and intuitively plausible - if hardly uncontroversial - criterion of moral rightness.¹⁵

According to this line of thinking, a moral statement is true if and only if it describes the world in the same way that a set of impartial observers in ideal conditions would. But the pronouncements of ideal observers are not what makes moral statements true. What **makes** a moral statement true is that there are **certain natural facts** that obtain now which determine what choices would be made under ideal conditions.¹⁶ At this point, the nature of the relationship between these natural facts and true moral **theories** remains crucially mysterious. The mystery centers on the indeterminacy of the ideal conditions which form the crucial link between them. The truth conditions for a moral theory, if there are any, involve reference to both natural facts **and** ideal conditions. As a result, if the ideal conditions for moral decision-making are indeterminate, so are the truth conditions for moral theories. Part of my goal, in the following sections of this chapter, will be to demonstrate that the ideal conditions for moral decision-making are indeed indeterminate. This will be

¹⁵Railton, 1986a, pp. 190,191.

¹⁶For a discussion of this point in the individual case see Railton, 1986a, p.174 and 1986b, p.25.

sufficient to show that the truth conditions for moral theories are also indeterminate. At any rate, at this point in the discussion it is far from clear that there is any set of natural conditions that **could** make a true moral theory true.

Now we can move on to the second question. Here, we want to know what it is about a true moral theory in virtue of which the moral facts it posits **must** obtain. Can we provide a characterization of the 'moral necessity' that holds between true moral theories and the moral facts they call for which is at least roughly analogous to the physical necessity that holds between true scientific theories and the scientific facts these theories call for? Again, Sturgeon gives us no answer to this question, even though, as we have seen, he is comfortable speaking in terms of the necessary truth of various moral statements. Following Railton's lead, the analysis would have to begin with the claim that it is not possible for a set of ideal observers to endorse a particular moral theory without at the same time espousing all of the particular moral judgments which that theory calls for. As a result, the natural facts that determine which moral theory would be endorsed under ideal conditions also determine which particular moral judgments would be endorsed under ideal conditions. If this strategy is successful, then, like the scientific case, there is a set of natural facts which both makes true moral theories true as well as makes it **necessary** that the particular moral judgments called for by true moral

theories are also true. It seems to me that a strategy like this cannot be successful because the same indeterminacy problems associated with ideal conditions that we saw in our discussion of the first question reappear here.

In order to explain why this is so, it is worth re-emphasizing that, according to naturalistic moral realism, the fact that a set of ideal judges in ideal conditions would espouse a particular set of moral judgments is **not** what **makes** these judgments true. Since moral facts are constituted by certain conglomerations of natural facts, the truth-makers for these moral judgments are the natural facts associated with each of them. These natural facts both make these judgments true as well as explain why the ideal observers espouse them. Now, we might wonder what role the ideal conditions are playing here if the truth conditions for moral statements can ultimately be specified in terms of natural facts about the world. The answer is that the ideal conditions play two distinct roles. On the one hand, the observers in ideal conditions are a heuristic device which enable us to discover moral truths. It is important to notice that they do not serve as a mechanism for generating them. They merely enable us to ascertain which conglomerations of natural facts are the moral ones. On the other hand, they **do** serve as a crucial part of the mechanism for generating true moral theories since they are **constructed** to be completely reliable detectors of moral facts. As such, they play a crucial role in actually

determining the nature of true moral theories. Since they play this crucial constitutive role, it is important that such ideal conditions be determinate. If they are not determinate, there **cannot** be a set of natural facts which both makes a true moral theory true as well as makes it the case that the particular moral facts called for by that theory **must** obtain.

We can bring the preceding discussion to bear directly on Sturgeon's explanation defense in the following manner. Recall that it is my position that there is something inherently problematic about referring to the moral realm as a factual one. According to Sturgeon, it is completely appropriate to refer to the moral realm as a factual one. He believes that scientific theories and moral theories are much more on a par than people think. One thing to notice is that the supposedly strong parallels between science and morality that form the backbone of the explanation defense begin to break down under closer scrutiny. We discover that Sturgeon has provided no positive characterization of what makes a moral theory true or even of what **could** make a moral theory true. Nor has he provided a positive characterization of what it is about a true moral theory in virtue of which the moral facts it posits **must** obtain. On the other hand, both of these characterizations have been provided fairly straightforwardly in the scientific case using the example of the proton in the cloud chamber. Keeping this in mind, it will be helpful to consider some of Sturgeon's comments concerning moral facts

and moral theories. After acknowledging that his account of the nature of moral facts is unclear at crucial points, he says:

A naturalist, however, will certainly want (and a critic of naturalism will likely demand) a fuller account than this of just where moral facts are supposed to fit in the natural world. ... The answer is that the account will have to be derived from our best moral theory, together with our best theory of the rest of the natural world - exactly as, for example, any reductive account of colors will have to be based on all we know about colors, including our best optical theory together with other parts of physics and perhaps psychology.¹⁷

He is well aware that these comments will not overwhelm the gainsayers. Further on, admits that:

Many philosophers will balk at this confident talk of our discovering some moral theory to be correct. But their objection is just the familiar one whose importance I acknowledged at the outset, before putting it to one side: For I grant that the difficulty we experience in settling moral issues, including issues in moral theory, is a problem (although perhaps not an insuperable one) for any version of moral realism. ... Our moral theory, if once we get it, will provide whatever reduction is to be had and will tell us where the moral facts fit.¹⁸

It seems to me that Sturgeon is simply talking through his hat in these passages. He speaks as if scientific and moral theories are of a piece when he has provided no sound theoretical basis for doing so. He has provided no positive account at all of the nature of the 'best moral theory'; yet, he speaks as if this notion is completely unproblematic. This was the point of bringing up Railton's discussion of ideal observers. It was merely a suggestion. It constitutes a way to

¹⁷Sturgeon, 1984, pp.60,61.

¹⁸Sturgeon, 1984, p.61.

flesh out the notion of 'best moral theory' naturalistically. The important similarity between Sturgeon's and Railton's approaches is that they both rely on an idealized framework to determine which sets of natural properties are the moral ones. For Sturgeon, the idealized framework is only vaguely cashed out as 'the best moral theory'; for Railton, it is characterized as a set of impartial observers in ideal conditions. If Sturgeon were to reject Railton's more detailed characterization of the best moral theory, the onus would be on him to produce some other characterization that remains within the limits of a naturalistic ontology. It is sufficient to note that the broad outlines of their approaches are basically the same. Thus, it is far from clear that Sturgeon has earned the right to speak of the moral realm as a factual one. It is far from clear that scientific theories and moral theories are of a piece, as Sturgeon presumes. Having brought a number of theoretical background issues out into the open, I will now proceed to argue that Sturgeon **cannot** plausibly speak of the moral realm as a factual one. Since he needs to be able to speak of the moral realm as a factual one in order for the explanation defense to successfully establish the existence of moral facts, showing that he cannot plausibly speak in this way will serve to defuse the explanation defense.

4.9: Strong Moral Properties

I will begin my case by describing an interesting point which was recently made by David Copp. He argues that the mere explanatory utility of a behavioral characteristic is not enough for it to qualify as a genuinely **moral** property. In order to see this, he claims,

we need only consider explanations that invoke concepts in moral perspectives we reject".¹⁹

Examples might be the Nietzschean concept of the 'overman' or Aristotelian character qualities.

A Nietzschean might seek to explain Stalin's ruthless behavior on the basis that he was an (approximation to the) overman. But this would be an ordinary psychological explanation that incidentally invoked a Nietzschean moral concept. ... [W]e would not regard the explanatory utility of the concept as justifying any standard that treats being overmanlike as a virtue or ideal.²⁰

For this reason, we would reject 'overmanness' as constituting a bona fide moral property even though it functions explanatorily. Part of the meaning of a legitimate **moral** predication involves the endorsement of the set of standards which that predication employs. In Copp's terminology, when we legitimately predicate a moral property of a particular state of affairs, we are making a 'strong' moral statement. We are

¹⁹Copp, 1990, p.247.

²⁰Copp, 1990, pp.247,248. Some might object that the discussion of the overman in this section is seriously misguided because it misconstrues Nietzsche's actual view. Copp's response in ftn.29 on p.247 is that "It is not relevant whether these remarks about Nietzsche are accurate." The discussion does not have to accurately portray Nietzsche's view in order to illustrate the point Copp wants to make here.

attributing a strong moral property to the state of affairs in question. If we are merely reporting that a particular state of affairs qualifies for the possession of a 'moral' property according to a set of standards which we don't necessarily endorse, then we are only making a 'weak' moral statement. We are only attributing a weak moral property to the state of affairs in question. When we do this, we are not making a genuine moral claim. Admitting that Stalin possessed the type of personality characteristics required to satisfy the conception of the overman does not commit one to the moral claim that Stalin possessed the moral property of overmanness.

According to this understanding of the nature of moral language use, the truth conditions for a strong moral proposition are two-faceted. Certain empirical conditions must hold and a certain moral standard must be justified. This is what the world must be like in order for a strong moral statement to be true. It must be a fact that certain empirical conditions obtain and it must be a fact that a certain moral standard is justified. It is crucial to notice that the moral standard's being justified is a necessary condition for the strong moral proposition's being true. If a person believes that only the empirical half of the truth conditions for a strong moral statement has been satisfied, then she will take that statement to be false. It is as if she is evaluating the truth of a conjunction where the truth conditions for only one of the conjuncts has been satisfied. The sincere utterer of a

strong moral statement such as 'Brutality is wrong' is committed to two propositions in Copp's scheme: (1) The practice of brutality contravenes a certain moral standard M. (2) The moral standard M is justified. Though moral standards are not propositional, statements about moral standards, such as (2), are. In Copp's words:

I propose that moral judgments are or entail propositions about putatively justified moral standards, such as a proposition to the effect that some moral standard is justified. This implies that a fully adequate account of the truth conditions of moral judgments would depend on an account of the conditions under which a moral standard would be justified.²¹

In this passage, Copp has pointed to a serious difficulty that the explanation defense has to face. For him, the problem with the explanation defense is that it fails to take account of an important element of moral language use. He claims that part of the meaning of a legitimate moral predication involves the endorsement of the set of standards which that predication employs. For him, this means that moral language use involves the predication of strong moral properties. It is interesting to notice that Copp's appeal to the notion of strong moral properties is intimately linked to the doctrine of internalism. According to Copp's characterization of moral language, assenting to the justification of the relevant moral standard is a necessary condition for the sincere assertion of a moral statement. Since assenting to the justification of a standard requires assenting to some reason to comply with the

²¹Copp, 1990, pp.254,255.

dictates of that standard, assenting to some reason to comply with its dictates is also a necessary condition for the sincere assertion of a moral statement. This is just to say that internalism must be true.

It is also interesting to notice that many readers have found Copp to be convincing on this point. This is an interesting **empirical** observation. It could have been the case that actual moral language users did not feel the force of Copp's objection at all. People could have said that Copp's objection makes no sense. But they don't. Copp's objection is convincing because he has put his finger on an inadequacy with the externalist characterization of moral language that is an implicit part of the explanation defense. This is just what we would expect if the a posteriori argument in favor of internalism in Chapter Two is correct.

The important point Copp makes here is that the actual nature of moral language use commits us to the position that if there are moral facts, then there are justified moral standards or theories. This is because the existence of a justified moral standard or theory is a necessary condition for the existence of a genuine moral fact. Justified moral standards or theories are constitutive aspects of genuine moral facts. As a result, if moral facts can be grounded naturalistically, then the justified moral standards or theories which constitute part of them can also be grounded naturalistically. Contrapositively, if justified moral

standards or theories **can't** be grounded naturalistically, then moral facts **can't** be grounded naturalistically either. Thus, demonstrating that justified moral standards or theories **can't** be grounded naturalistically will be sufficient to show that moral facts **can't** be grounded naturalistically. In this way, Copp's criticism can be used to argue that moral facts **can't** be grounded naturalistically.

This has important implications for the explanation defense. The cogency of the explanation defense is contingent on the assumption that moral facts **can** be grounded naturalistically. According to Sturgeon, it is legitimate to parallel moral inquiry and scientific inquiry precisely because the objects of moral inquiry are no different in kind from the objects of scientific inquiry. They are both varieties of natural facts. As Sturgeon admits:

As a philosophical naturalist, I take natural facts to be the only facts there are. If I am prepared to recognize moral facts, therefore, I must take them, too, to be natural facts:²²

Since the cogency of the explanation defense is contingent on the assumption that moral facts **can** be grounded naturalistically, Copp's criticism can be used as a basis for undermining the explanation defense in a way that he did not intend. Copp sees that there is a problem with the explanation defense as it stands, but he thinks that this problem only calls for a refinement within moral realism. He doesn't

²²Sturgeon, 1984, p.58.

anticipate that this problem could spell the end for moral realism. All that remains then, is to explain why justified moral standards or theories cannot be grounded naturalistically. This will show that no natural set of features could satisfy the required truth conditions for strong moral statements. This will show that there cannot be any naturalistic moral facts.

4.10: Justified Standards And Ideal Conditions

A standard way to formulate justified moral standards is by attempting to determine the decisions that would be made under some set of **ideal conditions**. Sturgeon is committed to some version of this strategy since he relies on the **best** moral theory to demarcate the extensions of moral properties. The strategy works in the following way. If an audience composed of beings very much like ourselves were to formulate a set of standards of conduct in an ideal epistemic situation, we could, by definition, stipulate that the set of standards so formulated would constitute a **justified** set of moral standards for us. They would be the standards that accrue from the best moral theory. In this approach, the problem of arriving at a justified set of moral standards is reduced to one of determining the nature of these ideal epistemic conditions. The problem, however, is that it might not be possible to provide an entirely naturalistic specification of the ideal epistemic conditions for moral decision-making. To be sure, we may be able to specify a number of uncontroversial

requirements for the ideal situation. For example, having reliable information about the empirical effects of adopting various policies might have to be part of any plausible characterization of such ideal conditions. But these few obvious requirements will not be sufficient to provide a substantial positive characterization of the ideal conditions for moral decision-making. It will still be a live question whether a number of other constituents should or should not be included. If this issue cannot be resolved, we cannot specify **the** ideal set of epistemic conditions that will produce a **justified** set of moral standards. For any substantial set of putatively ideal conditions, we can question whether or not we would be justified in employing that set of 'ideal' conditions as a moral decision-making matrix as opposed to another one. If it can't be shown to be justified in a non-circular manner, then the moral standards which accrue from that set of 'ideal' conditions will also lose their status as justified.

In what follows, I argue that there is no genuinely **ideal** moral decision-making matrix which is specified by the naturalistic fabric of the universe. As a result, justified moral standards or theories cannot be grounded naturalistically. As a result, there are no naturalistic moral facts. Thus, the explanation defense fails. I will use a concrete example to illustrate the point I want to make here. I am using it as a means of illustrating what would have to be the case for a strong moral statement to be true. My intention

is to convince the reader that no natural set of features could satisfy the required truth conditions. From there, I will attempt to convince the reader that the same difficulty arises in the moral scenarios that Sturgeon discusses. Here is the example.

4.11: Tar-Maintaining

Suppose you are a member of a series of anthropological teams engaged in the large-scale project of recording the practices of indigenous cultures in Africa. Your team has been assigned to study a farming people in Nigeria called the Tiv. Though the team is initially shunned by the community, you are able to earn their acceptance after several months of living amongst them. Eventually, an elderly tribesman becomes your confidant and explains to you the significance of many of the goings-on in the day to day life of the community. One thing you begin to discern is that a peculiar cliquishness pervades the affairs of the community. This is a reflection of the community's social structure.

The Tiv are organized in family compounds consisting primarily of brothers, their wives, and their children. These groups are distributed more or less uniformly over a district of patrilineally related kin, known as a **tar**.²³

On one occasion, you notice a young man carrying a large bag of grain coming out of the hut of a local family which you know to belong to a rival clan. A number of his fellow clan members also witness this event, but pretend not to notice

²³Miller, 1992, p.22.

him. Upon discovering his bag of grain missing, the enraged owner accuses the young man of the theft and demands that the matter be settled before the local court. Apparently, the two had been involved in a bitter argument the previous day, which ended with threats of revenge by the young man. At the trial, you are astonished to hear a parade of the young man's relatives testify that they had not seen him anywhere near the plaintiff's house that day because you can remember the young man walking right past a number of these witnesses with the bag of grain in hand. You feel incensed at the travesty of justice taking place before you but your confidant convinces you that it is not your place to speak out here. The trial unfolds and, to your dismay, a judgment is made in favor of the accused. The judgment is accepted by all and the matter is resolved. The next day while walking with your confidant you notice one of the witnesses who lied at the trial sitting in front of her hut. Out of curiosity, you ask him why she lied at the trial. You are told that she testified in the way she did because she is a **tar-maintainer**. We might say that the explanation of her acting the way she did is that she possesses a deed-causing moral property which causally influenced her behavior. This all seems perplexing until you come to understand the nature of Tiv morality more fully. It turns out that:

The central moral problem of the Tiv is to protect the network of relationships from consequent rupture without unnecessary harm to individuality and self-esteem. As the Tiv sometimes put it, when conflicts threaten to "spoil

the **tar**," there is an urgent need to bring disputants together and "repair the **tar**."

The main means of repairing the tar is organized discussion, consultation, social pressure, and occasionally, coercion in the **jir**, the "native court" in colonial jargon.²⁴

This has important social implications for situations like the preceding courtroom scenario. The significance of the woman's lying can only be adequately understood in terms of the network of social obligations which are such a crucial part of Tiv society. These play a crucial role in determining the nature of appropriate courtroom behavior.

Except for police, witnesses are, almost exclusively, brought to court by individual disputants and are of two categories: relatives, who are expected to testify in the interests of their disputant, and witnesses preselected by the disputant and paid for their testimony. ... As a result, there is much false testimony by relatives, and the burden of finding and negotiating with witnesses keeps much truth out of court. Yet, for the Tiv, the witness who freely steps forward and gives incriminating evidence does wrong if, as is likely, he or she "spoils the **tar**."²⁵

The reason that the woman lied in court then, is that she was predisposed to comply with social pressures to minimize tension within her own social unit even if this meant testifying to falsehoods.

The question I wish to focus in on here is whether or not this woman can be adequately characterized as possessing a particular **moral property** which explains why she acted in the manner she did. According to the preceding analysis, it can

²⁴Miller, 1992, p.22.

²⁵Miller, 1992, p.23.

only be a moral property if a moral standard advocating the practice of tar-maintaining is justified. From here, we can consider the concrete issue of what would have to be the case for it to be true that the practice of tar-maintaining is justified. As I have noted, the standard answer is that tar-maintaining is justified if an adequately representative group of impartial observers would deem the practice to be acceptable under ideal epistemic circumstances.

4.12: Normative Presuppositions And Ideal Situations

At this point, we can consider some comments made by Rawls on what the ideal decision-making procedure really amounts to. He claims that:

What justifies a conception of justice is not its being true to an order antecedent to and given to us, but its congruence with our deeper understanding of ourselves and our aspirations, and our realization that, given our history and the traditions embedded in our public life, it is the most reasonable doctrine for us. ... Apart from the procedure of constructing the principles of justice there are no moral facts. Whether certain facts are to be recognized as reasons of right and justice, or how much they are to count, can be ascertained only from within the constructive procedure...²⁶

Rawls' comments on justice can be generalized to apply to the broader notion of morality itself. His claim is that we seriously misunderstand the nature of moral decision-making procedures if we look at them as reliable mechanisms for detecting moral facts. The reason for this is that there are no moral facts there to be detected which exist independently of the decision-making procedure which is supposedly doing the

²⁶Rawls, 1980, p.519.

detecting. According to Rawls, given that we are the sorts of creatures that we are, living in the circumstances in which we find ourselves, the ideal conditions provide a matrix which will allow us to arrive at the best set of standards of conduct for us.

A problem with this kind of setup, however, is that the set of dispositions and desires which some or all of us possess could be morally suspect. As a result, even the standards of conduct which we would espouse under ideal conditions would be morally suspect. There is no guarantee that mere placing of a person or group of persons in the ideal circumstances for moral decision-making will automatically expunge all morally questionable dispositions from her. For example, Susan Babbitt describes how a person like Thomas Hill's 'Deferential Wife' would fare in such ideal circumstances.

Equipped with ideal cognitive capacities and resources, it is not clear that a person who is degraded and diminished by social conditioning would have reason to choose goods typically thought to represent flourishing.²⁷

This problem could be remedied by describing the ideal decision-making matrix in such a way that all of the participants have undergone 'transformation experiences' which have excised any morally unwholesome personality traits. In the case of the Deferential Wife,

²⁷Babbitt, 1993, pp.248,49.

In order for it to be rational for her to desire autonomy in the sense that rules out her habitual servility, her actual sense of self would have to be transformed so that habitual servility is not what defines it. But defining a person's objective interest in terms of a perspective the person might have but in fact does not is just what the liberal view rules out.²⁸

It strikes me that Babbitt's point against Rawls is well taken. It also seems to me that this point can be applied to the moral realism debate in a way that Babbitt did not have in mind when she formulated her case against Rawls.

The problem is that there are **normative** constraints on what many of us will count as a genuinely **ideal** normative decision making matrix. If this is so, then the ideal decision-making matrix can hardly serve as a source of justification for a normative stance since the validity of certain normative claims is presupposed in the very character of the ideal conditions. This is what **makes** them ideal conditions in the first place. One must already know what the 'correct' normative perspective is in order to be able to determine the nature of genuinely ideal conditions for normative decision-making. Determining what is going to count as **the** set of ideal moral decision-making conditions is not a normatively neutral matter. Various candidate proposals for ideal conditions will be initially assessed in terms of their normative consequences. Those that don't yield the 'right' normative results will be screened out in advance.

²⁸Babbitt, 1993, p.250.

Consider Babbitt's criticism of Rawls in this light. The problem she has with the decisions made by the Deferential Wife in Rawls' proposed set of ideal epistemic conditions is simply that they are **normatively unacceptable**. Babbitt attempts to explain the difficulty here by pointing to the fact that the Deferential Wife has a diminished sense of self which prevents her from making the right normative decisions even when she is in a supposedly optimal epistemic situation. But it seems to me that nothing is gained by attempting to explain her decisions in terms of her lacking an 'adequate sense of self'. All that having an adequate sense of self **means** here is having the dispositions to espouse certain normative judgments and to reject others. If we attempt to justify moral standards in terms of ideal conditions, we are faced with the following question. Why are so many of us unwilling just to say that if a life of perpetual servility and deference to her cherished husband is the kind of life she would opt for in ideal epistemic conditions, then that kind of life must **really** be what is good for her? If there really are objective moral facts that can be detected by the heuristic device of ideal conditions, it might be that the ideal conditions reveal surprising results in cases like hers. How can we be certain about what the world is like in advance of using ideal detecting mechanisms to discover what it is like? The set of moral properties out there to be discovered might not have quite the character we initially envisaged.

But many of us are not open to this possibility. Instead, we are convinced that any conditions which would yield these kinds of decisions cannot, **for that very reason**, be genuinely ideal. Whether or not a particular set of conditions is ideal, then, cannot be determined independently from and in advance of the set of normative judgments that the set of conditions ultimately yields. As a result, the justificatory status given to the set judgments made under a particular set of 'ideal' conditions ultimately accrues from the normative presuppositions which initially determine the character of such ideal conditions. In judging a set of circumstances for moral decision-making to be an ideal one, we are conferring an honorific status upon this set of circumstances. We are claiming that it is able to yield standards which are truly justified. But if this honorific status is contingent on the circumstances generating the **right** sort of normative judgments, then the judgments it yields are justified only if the normative standards which conferred the status of 'ideal' on the circumstances in the first place are themselves justified. If this is so, then the circumstances' being ideal is not the ultimate source of normative justification for the normative judgments which are generated in them. The ultimate source of justification for the normative judgments generated in the ideal circumstances must be the same as the source of justification for the normative judgments which initially determine the nature of such circumstances. The problem is

that these latter judgments have no source of justification. They simply fall into the realm of the 'intuitively obvious'.

Our reaction to Babbitt's diagnosis of the Deferential Housewife is a case in point. Many of us take the Deferential Housewife's reluctance to concur with our judgments in matters of ultimate moral importance as an indication that there is something wrong with her. To not concur with such judgments is in some way to have a morally flawed character. It is to have a disposition to act in ways which will arouse the contempt or pity of 'normal', 'ordinary' people. As a society, we have a more or less homogeneous corporate disposition to value certain types of behavior and shun other types. This is part of our form of life. We have shared moral ground on certain moral questions. Certain attitudinal biases are built in to the way that many of us view moral decision-making. They are also built into the way that many of us evaluate personality types or life aspirations. Any 'plausible' metaethical position has to conform to the dictates of these biases. As a result, any plausible set of ideal circumstances for moral decision-making has to conform to the dictates of these biases.

4.13: Tar-Maintaining Reconsidered

We can now return to the Tiv example discussed earlier. Recall that the problem we are considering is whether or not the woman's propensity to tar-maintain is adequately characterized as the possession of a moral property. In order

for it to be the case that the disposition to tar-maintain is a moral property, it must be the case that a moral standard that endorses or condemns the practice of tar-maintaining is justified. In order for that to be the case, it must be the case that tar-maintaining would be endorsed under optimal circumstances. The problem, however, is that the average Tiv and the average North American would be unable to come to any consensus on the nature of such optimal circumstances. The reason for this is that each of them would evaluate the various candidate proposals for ideal circumstances in terms of his own normative biases. An important difference between the average North American and the average Tiv is that the former is predisposed to value abstractness and impartiality in moral decision-making while the latter is not. As a result, abstractness and impartiality will be crucial features in the North American's optimal circumstances but not in the Tiv's. It should be emphasized that the disagreement between the North American and the Tiv cannot be adequately characterized as a **factual** one. Each of them could be fully aware that the practice of tar-maintaining would be rejected in the North American's optimal circumstances and accepted in the Tiv's optimal circumstances. The root of their deep-seated disagreement lies in their differing UMPs. As a result of their differing UMPs, each of the two parties has a different view of the nature of optimal circumstances for moral decision-making.

The problem is that there is no neutral UMP we can turn to in order to assess the relative merits of the two competing characterizations of 'optimal'. There is no neutral set of criteria which can be used to arrive at the truly optimal set of circumstances for moral decision-making. A feature like impartiality only becomes a requirement for the optimal circumstances if we assume that the individualistic perspective on moral matters that the North American is inclined to endorse is essentially correct. If one is not already inclined to endorse this kind of UMP, then the incorporation of impartiality into a moral decision-making matrix will not count as a point in its favor. The Tiv would not count impartiality as a virtue of a moral decision-making matrix precisely because he endorses an UMP on moral matters which places ultimate value on the fulfilment of social roles. The criticisms the North American might make of his group-based morality would not count as criticisms to him since he does not hold the kind of individualistic UMP that is presupposed by such criticisms. In the same way, the criticisms the Tiv might make of the North American's individualistic morality would not count as criticisms to the North American since he does not endorse the kind of corporate UMP presupposed by these criticisms. At the end of the day, neither of the two camps will be able to convince the other of the superiority of his UMP without begging the question. Notice that their deep-seated moral disagreement cannot be

resolved by simply pointing out that each of the two parties is advocating the employment of a different set of general means to achieve the same end of social harmony. The problem is that the two parties have a fundamental moral disagreement over whether or not the practice of lying should be part of this general means. This is the issue they cannot see eye-to-eye on. Due to their differing deep-seated moral impulses, each will continue to endorse his own view of the situation. In order to change a person's attitudes about the nature of optimal circumstances for moral decision-making, it will be necessary to change his deep-seated normative attitudes about the kind of lives we should be living. It will be necessary to change his UMP.

Thus, there is no objective way to adjudicate between the competing characterizations of ideal circumstances for moral decision-making endorsed by each of the two parties. There are no natural features of the world that we can point to in order to resolve this kind of dispute. There are no natural features of the world that will ground a particular set of ideal circumstances for moral decision-making. Since there is no way to ground a particular set of ideal circumstances for moral decision-making naturalistically, there is no way to justify moral standards naturalistically. Since justified moral standards are a constitutive aspect of moral facts, it is not possible to ground moral facts naturalistically. Thus, the explanation defense fails. The explanation defense fails

because it **assumes** that moral facts can be grounded naturalistically.

4.14: Irreconcilable Differences

We can now consider Sturgeon's claim that 'It is a fact that Hitler was morally depraved' more closely. In doing so, we will see that the problem of deep-seated irreconcilable differences also arises here. We need to be aware that Sturgeon's comments about Hitler take place within a shared form of life. The fact that the average reader shares the same attitudinal biases as Sturgeon explains why so many of us are ready to endorse the judgment that Hitler was morally depraved. It is part of the form of life that we share to view a person like Hitler with contempt. It is not necessary to present a case for the validity of the judgment that Hitler was morally depraved. Such a judgment already resonates well with the moral intuitions many of us hold about the way that people should be. There are no lingering doubts about whether or not we should endorse such a judgment which need to be assuaged by rational argumentation. But this unconditional willingness on our part to endorse the judgment that Hitler was morally depraved in no way makes such a judgment a **factual** one. Since many of us are in hearty **normative** agreement with Sturgeon on this issue, the **metaethical** issue of whether such a statement is factual or not seems rather moot. Unfortunately, there are people in this world who are not in normative agreement with Sturgeon's assessment of Hitler's

character. According to many neo-Nazis, Hitler was a great man whose only shortcoming was his inability to succeed at bringing his grand vision to fruition. These people don't believe that Hitler was morally depraved because they don't believe that there is a justified moral standard which censures a character like Hitler's. They don't believe this because of the deep-seated moral intuitions they possess. They partake of a form of life which is radically different from our own.

If the moral realm is a factual one, then the problem with such people is simply that they hold false beliefs. This strikes me as an implausible characterization of the situation, however. What many of us find repulsive about these people is not so much a matter of what they believe as a matter of the kind of people that they are. They are the kind of people who are willing to endorse the actions of a person like Hitler because they don't take the criticisms that the rest of us make about his life to be legitimate criticisms. They have a different conception of optimal moral circumstances which embodies a different set of relevant criteria for reliable moral decision-making. As a result, they are apt to endorse the same kinds of things that Hitler would endorse. Like Hitler, they would claim that it is people like us who are the morally bankrupt ones. It is naive to suppose that these people are merely factually misinformed about certain issues. It's just not true that we can bring them

around to our way of thinking simply by apprising them of a number of what we take to be important facts of the case. Unless one is already of the view that these kinds of considerations matter morally, being reminded of them will have no effect on one's normative outlook.

At this point, we need to look at the kind of UMP I am discussing a little more closely. We can begin by considering the following aphorism by Nietzsche:

Who can attain to anything great if he does not feel in himself the force and will to **inflict** great pain? The ability to suffer is a small matter: in that line, weak women and even slaves often attain masterliness. But not to perish from internal distress and doubt when one inflicts great suffering and hears the cry of it - that is great, that belongs to greatness.²⁹

The presupposition that certain kinds of lives have more value than other kinds of lives is a crucial component of this type of UMP. According to this point of view, the world is a better place when the most valuable kinds of lives are given priority than it is when every kind of life is afforded equal consideration. It is not important to convince those with less valuable lives of the legitimacy of this perspective. As Nietzsche claims:

The essential thing in a good and healthy aristocracy is, however, that it does **not** feel itself to be a function (of the monarchy or of the commonwealth) but as their **meaning** and supreme justification - that it therefore accepts with a good conscience the sacrifice of innumerable men who **for its sake** have to be suppressed and reduced to imperfect men, to slaves and instruments. Its fundamental faith must be that society should **not** exist for the sake of society but only as a foundation

²⁹Brown, 1968, p.140.

and scaffolding upon which a select species of being is able to raise itself to its higher task and in general to a higher **existence**...³⁰

According to this kind of UMP, those who espouse egalitarian doctrines are actually the morally depraved ones because egalitarian social policies will only result in the spread of decadence throughout our society. The potentially great will be brought down in order that the mediocre might prosper. Society will be brought down to the level of the lowest common denominator. It is a natural facet of our earthly existence that the base majority must suffer and make sacrifices in order that the great minority might have the opportunity to flourish. It is a sign of moral weakness to be moved by this suffering. The morally virtuous person will work to bring about the flourishing of the elite even if the masses must suffer in the process. In fact, he will revel in their suffering.

The magnitude of an "advance" can even be measured by the mass of things that had to be sacrificed to it; mankind in the mass sacrificed to the prosperity of a single **stronger** species of man - that **would** be an advance.³¹

The morally weak person will allow this suffering to draw him away from the important task of promoting greatness. He will begin to give the mediocre masses more consideration than they ought to have. Once the interests of the masses are counted on a par with the interests of the elite, decadence can be the

³⁰Nietzsche, 1886, p.174.

³¹Nietzsche, 1887, p.78.

only result. To count the interests of all on a par is the essence of moral depravity. According to this kind of UMP, it is not morally appropriate to count the interests of everyone on a par. Thus, the fact that a particular policy is inimical to the interests of certain people and does not give equal consideration to their point of view does not count as a moral reason to reject that policy. If the adoption of that policy is conducive to the achievement of greatness by the few, then it is a good policy on those grounds alone. It is a necessary means to a laudable end. In cases like these, the end does justify the means. The morally upright person will recognize this and will not allow himself to become squeamish and neglect to do what needs to be done.

A person who endorsed this kind of UMP would deny the claim that Hitler was morally depraved.³² According to the 'Nietzschean', the worst that could be said about Hitler is that he was factually misinformed about the nature of true greatness. There is no reason to think that all of those who are potentially great will be confined to the Aryan race. If Hitler had been acting in the interests of the right kind of people, however, there would be nothing wrong with the things he had done. The wrongness of his actions, then, can be traced

³²Like Copp, I am prepared to leave the question of whether or not the preceding characterization accurately represents a balanced picture of Nietzsche's view to the side. In what follows, I will refer to the view as 'Nietzschean'. The scarequotes will indicate that I am not necessarily attributing the viewpoint I am discussing to Nietzsche himself.

to a number of mistaken beliefs rather than to depravity of character.

4.15: Morally Relevant Differences

One might respond to the preceding description by claiming that the 'Nietzschean' needn't be taken seriously since he obviously isn't a **competent** moral judge. If we don't want to allow the possibility that this kind of person is morally competent, though, we should be able to explicate the basis of our incompetence claim. The essence of the incompetence claim would be that the 'Nietzschean' is unable to distinguish between morally relevant and morally irrelevant differences. The notion of a morally relevant difference is closely connected to the presupposition of **consistency** in moral theorizing. If we assume that morally acceptable conduct must be consistent, then in our moral dealings with others we are required to treat like cases alike. We can justify treating two similar cases differently only by pointing to a morally relevant difference between them. On this basis, we can claim that the two cases are not alike in a respect which is important enough to merit differential treatment of the two cases. A theory of what constitutes a morally relevant difference underlies such a claim, however. This theory will specify that certain kinds of differences merit differential treatment while other kinds of differences do not.

The dispute that many of us have with the 'Nietzschean' can be characterized as a disagreement over what counts as a

morally relevant difference. We believe that the fact that someone has a potential for a certain kind of greatness is not a morally relevant difference which entitles him to differential treatment. The 'Nietzschean' believes that it is. In order to resolve such a dispute, we would have to find some grounds for claiming that one way of parsing out morally relevant differences is superior to another. We would have to be able to substantiate the claim that one kind of UMP is able to zero in on the differences which are morally relevant while the other is not. In order for this to make sense, it must be the case that some differences **really are** morally relevant while others are not.

Since the moral realist endorses the existence of objective moral facts, she is committed to defending the claim that there really are morally relevant differences which are there to be discovered. According to her, the problem with the 'Nietzschean' is that he is in some sense morally blind. He is unable to make moral differentiations in the way that a morally healthy person does because he perceives the world through an inappropriate set of categories. His problem is analogous to the problem that a color blind person has. Because there is something wrong with him, he is unable to carve up the world correctly. Some differences which are morally relevant, he takes to be morally insignificant; other differences which are morally insignificant, he takes to be

morally relevant. He is out of touch with the way that the world is.

In order to substantiate this description of the 'Nietzschean', the moral realist must be able to provide a naturalistic grounding for the claim that certain differences really are morally relevant while other differences are not. If indeed this is something which is out there to be discovered, then there must be some naturalistic basis which underlies it. Something in the world must make it the case that basing our moral decision-making on certain kinds of differences is correct because they **are** the morally relevant ones while basing our moral decision-making on other kinds of differences is incorrect because such differences **aren't** morally relevant ones. If the moral realist is unable to furnish grounds for the claim that certain types of differences really are morally relevant while others are not, then there seems to be no basis for the claim that strong moral properties are part of the natural furniture of the world. All we have are different moral schemes which base moral decision-making on the presence or absence of various naturalistic criteria. The world does not require that genuine moral decision-making be conducted on the basis of certain naturalistic criteria and not others.

Bruce Brower attempts to provide a positive characterization of what it is in the world which makes it the case that certain differences are morally relevant while

others are not. For him, the crucial state of affairs is our **actual dispositions** to respond to certain states of affairs in certain ways and other states of affairs in other ways. In his words:

Dispositional ethical realism is the view that ethical properties are specified by empirically discoverable, reductive accounts that treat the moral properties as constitutively dependent on evaluators' responses or dispositions to respond. Any such account can be represented by a biconditional of this form:

1. X has value V if and only if evaluators would, under appropriate conditions, respond to X with reaction R.³³

Rather than looking completely outside of ourselves for the truth conditions for moral statements, we can include our own dispositions to respond as part of these truth conditions. The differences which are morally relevant, then, are the ones which we have a disposition to count as morally relevant. That we have a disposition to count them as morally relevant is what **makes** them morally relevant. It is important to notice that the 'under appropriate conditions' clause plays a crucial role in this account of moral properties. Due to actual circumstances of life, people are often biased regarding what they will count as morally relevant. Our actual dispositions to respond must be purified from these vitiating influences if they are to constitute a legitimate basis for the determination of morally relevant differences. This can be done by imagining what our actual dispositions to respond would be if we were placed in an **idealized** kind of situation.

³³Brower, 1993, p.222.

There, our dispositions would remain the same, but the vitiating influences would be removed. This would allow us to isolate our best responses to moral situations. These could then constitute an acceptable naturalistic basis for the determination of morally relevant differences. The problem with this move, of course, is that unless we can provide a naturalistic definition of an idealized situation, we cannot provide a naturalistic reduction that would determine which of the many possible responses our actual dispositions would endorse in various circumstances are our **best** responses. There is a metaphysical cost involved in the purification of our actual dispositions to respond to various moral phenomena. The cost is that our purified dispositions are no longer part of the furniture of the world. As such, they cannot count as the naturalistic truth-conditions which determine which differences in the world are morally relevant and which are not. They are only theoretical constructs whose nature will vary according to the particular conception of idealized conditions one espouses. We can even characterize the disagreement between the 'Nietzschean' and many of us as one over the nature of the conditions which will result in our **best** responses. There are no naturalistic features of the world which can solve this dispute because sets of naturalistic features aren't the kind of things which can resolve this dispute.

4.16: Concluding Remarks

I will now summarize my critique of the explanation defense. The problem with calling a statement like 'Hitler was morally depraved' true is that there is no naturalistic state of affairs available that can make this statement true. There is no fact of the matter about Hitler which obtains independently of endorsing an evaluative perspective on the nature of appropriate human conduct which is able to make that statement true in the strong sense. Someone might object that any **plausible** perspective on the way that humans should conduct themselves would have to endorse that statement. I agree. My consensus isn't going to help here, unfortunately, since it merely betrays the fact that I embrace an UMP that resonates with hers on this issue. The important metaethical issue is whether or not this **inevitability of endorsement** within our overlapping UMPs is somehow able to make the statement we endorse a factual one which admits of truth or falsity. How does the fact that it is not plausible to withhold endorsement of that statement within our UMPs **make** the statement itself pertain to a matter of fact?

Suppose someone claimed that it did. The presupposition underlying this claim would be that the truth conditions for the statement 'Hitler was morally depraved' are constituted by the fact that the statement would be endorsed by all 'plausible' UMPs. If we want to maintain that the statement 'Hitler was morally depraved' is true on the grounds that it

would be endorsed by all plausible UMPs, however, we will need some naturalistic criteria for distinguishing plausible from implausible UMPs. Some objective features of the world must make it the case that certain UMPs are plausible while others are not. If we cannot point to such objective features of the world, then we will merely be preaching to the converted in our normative argumentation. It is not surprising that someone who is predisposed to endorse 'Hitler was morally depraved' will take any UMP which fails to endorse 'Hitler was morally depraved' to be implausible. The crucial question is what **makes** such an UMP implausible apart from the normatively unpalatable judgments it espouses? If no answer can be given to this question, then the moral realist has no metaethical grounds for referring to 'Hitler was morally depraved' as true.

It may be disconcerting to some to countenance the possibility that there are no objective features of the world which make it the case that 'Hitler was morally depraved' is true. There is an irresistible urge to claim that we are **objectively right** and the neo-Nazis are **objectively wrong** in our respective assessments of Hitler's character. If we look at matters dispassionately, however, we can see that it is by no means clear how any positive account of the truth conditions which would make us objectively right and the neo-Nazis objectively wrong can be provided. There is no reason that this realization should make us any less vigilant in our

normative stance, however. Many of us have become the kind of creatures who cannot but find a person like Hitler to be morally repulsive. To deny this would be to deny what we are. It would be to deny what we take to be important and valuable at an ultimate level. It could have been the case that our UMPs were substantially different, but, in fact, they aren't. As human beings, we must make moral choices in the situation in which we find ourselves. There is no point in positing features of the world in the hope of providing some ultimate metaphysical grounding for such choices if no rationally compelling argument for the existence of just the kind of features necessary to do this job can be provided.

Bibliography

- Babbitt, Susan E. 1993. "Feminism And Objective Interests: The Role Of Transformation Experiences In Rational Deliberation." pp. 245-264 in Linda Alcoff and Elizabeth Potter, eds., **Feminist Epistemologies**. London: Routledge.
- Blackburn, Simon. 1984. **Spreading The Word**. Oxford: Clarendon Press.
- Boyd, Richard N. 1983. "How To Be A Moral Realist." pp. 181-228 in Sayre-McCord.
- Bright, Robert. 1991. **Foundations Of Utilitarianism**. Unpublished Doctoral Dissertation, Dalhousie University: Halifax, Nova Scotia.
- Brink, David O. 1984. "Moral Realism And The Sceptical Arguments From Disagreement And Queerness." **Australasian Journal Of Philosophy** 62:111-125.
- . 1986. "Utilitarian Morality And The Personal Point Of View." **The Journal Of Philosophy** 83:417-438.
- . 1989. **Moral Realism And The Foundations Of Ethics**. Cambridge: Cambridge University Press.
- Brower, Bruce W. 1993. "Dispositional Ethical Realism." **Ethics** 103:221-249.
- Brown, Colin. 1968. **Philosophy And The Christian Faith**. Downers Grove, Ill.: InterVarsity Press.
- Burge, Tyler. 1979. "Individualism And The Mental." **Midwest Studies In Philosophy** 4:73-121.
- Campbell, Richmond. 1993. "Critical Notice: Allan Gibbard, *Wise Choices, Apt Feelings: A Theory Of Normative Judgment*." **Canadian Journal Of Philosophy** 23:299-324.
- Copp, David. 1990. "Explanation And Justification In Ethics." **Ethics** 100:237-258.
- Darwell, Stephen L., Allan Gibbard, and Peter Railton. 1992. "Toward Fin de siecle Ethics: Some Trends." **The Philosophical Review** 101:115-190.
- DePaul, Michael R. 1993. "Critical Notice: Moral Realism And The Foundations Of Ethics." **Philosophy And Phenomenological Research** 53:531-535.

- Frankena, W.K. 1939. "The Naturalistic Fallacy." *Mind* 48:464-477.
- Geach, Peter. 1965. "Assertion." *Philosophical Review* 74:449-465.
- Gibbard, Allan. 1990. *Wise Choices, Apt Feelings*. Cambridge, Mass.: Harvard University Press.
- Goldsworthy, Jeffrey. 1992. "Externalism, Internalism, And Moral Scepticism." *Australasian Journal Of Philosophy* 70:40-60.
- Hare, R.M. 1952. *The Language Of Morals*. Oxford: Clarendon Press.
- . 1963. *Freedom And Reason*. Oxford: Clarendon Press.
- Harman, Gilbert. 1965. "Inference To The Best Explanation." *The Philosophical Review* 74:88-95.
- . 1977. *The Nature Of Morality: An Introduction To Ethics*. New York: Oxford University Press.
- Horgan, Terence and Timmons, Mark. 1990. "New Wave Moral Realism Meets Moral Twin Earth" *Journal Of Philosophical Research* 16:447-472.
- Jobe, Evan K. 1990. "Sturgeon's Defense Of Moral Realism." *Dialogue* 29:267-275.
- Kornblith, Hilary. 1985. "Beyond Foundationalism and Coherence Theory." pp.115-127 in Hilary Kornblith, ed., *Naturalizing Epistemology*. Cambridge, Mass.: MIT Press.
- Kripke, Saul. 1972. *Naming And Necessity*. Cambridge Mass.: Harvard University Press.
- Little, Margaret. 1994a. "Moral Realism I: Naturalism." *Philosophical Books* 35:145-153.
- . 1994b. "Moral Realism II: Non-Naturalism." *Philosophical Books* 35:225-233.
- Mackie, J.L. 1977. *Ethics: Inventing Right And Wrong*. New York: Penguin.
- McDowell, John. 1985. "Values And Secondary Qualities." pp. 166-180 in Sayre-McCord.
- McNaughton, David. 1988. *Moral Vision: An Introduction To Ethics*. New York: Basil Blackwell.

- Miller, Richard W. 1992. **Moral Differences**. Princeton: Princeton University Press.
- Moore, George Edward. 1903. **Principia Ethica**. Cambridge: Cambridge University Press, 1948.
- . 1910. **Some Main Problems Of Philosophy**. New York: MacMillan, 1953.
- Nietzsche, Friedrich. 1886. **Beyond Good And Evil**. R.J. Hollingdale Tr., Markham, Ont.: Penguin Books, 1984.
- , 1887. **On The Genealogy Of Morals**. Walter Kaufman Tr., New York: Random House, 1969.
- Putnam, Hilary. 1973. "Meaning And Reference." pp. 308-315 in A.P. Martinich, ed., **The Philosophy Of Language**. New York: Oxford University Press, 1990.
- . 1981. **Reason, Truth, And History**. Cambridge: Cambridge University Press.
- Railton, Peter. 1986a. "Moral Realism." **The Philosophical Review** 95:163-206.
- . 1986b. "Facts And Values." **Philosophical Topics** 14:5-31.
- . 1989. "Naturalism And Prescriptivity." **Social Philosophy And Policy** 7:151-174.
- Rawls, John. 1980. "Kantian Constructivism In Moral Theory." **Journal Of Philosophy** 77:515-572.
- Raz, Joseph. 1979. **The Authority Of Law**. Oxford: Clarendon Press.
- Regan, Tom. 1986. **Bloomsbury's Prophet: G.E. Moore And The Development Of His Moral Philosophy**. Philadelphia: Temple University Press.
- Sayre-McCord, Geoffrey. 1988. **Essays On Moral Realism**. Ithica: Cornell University Press.
- . 1988. "Moral Theory And Explanatory Impotence." pp.256-281 in Sayre-McCord.
- Scheffler, Samuel. 1982. **The Rejection Of Consequentialism**. Oxford: Clarendon Press.
- Schotch, Peter K. 1991. "Chapter 2: Modal Logic." Unpublished Manuscript.

- . 1992. "Remarks On 'Copenhagen Semantics'" Unpublished Manuscript.
- Searle, John R. 1962. "Meaning And Speech Acts." **Philosophical Review** 71: 423-432.
- Solomon, Wm. David. 1987. "Moral Realism And The Amoralist." **Midwest Studies In Philosophy** 12:377-392.
- Sturgeon, Nicholas. 1984. "Moral Explanations." pp. 49-78 in David Copp and David Zimmerman, eds., **Morality, Reason, And Truth**. Totowa, N.J.: Rowman and Allenheld.
- . 1986. "What Difference Does It Make Whether Moral Realism Is True?" **The Southern Journal Of Philosophy, Supplement** 24:115-141.
- White, Alan R. 1958. **G.E. Moore: A Critical Exposition**. Oxford: Basil Blackwell.
- Wiggins, David. 1987. "Truth, Invention, And The Meaning Of Life." pp. 127-165 in Sayre McCord.
- Williams, Bernard. 1973. "A Critique Of Utilitarianism." pp.75-150 in J.J.C. Smart and Bernard Williams, **Utilitarianism: For And Against**. Cambridge: Cambridge University Press.