



THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

Research Commons

<http://waikato.researchgateway.ac.nz/>

Research Commons at the University of Waikato

Copyright Statement:

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

The thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author's right to be identified as the author of the thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from the thesis.

VIRTUE AND SELF-INTEREST

A thesis
submitted in fulfilment
of the requirements for the degree
of
Doctor of Philosophy in Philosophy
at
The University of Waikato
by
TERY HARDWICKE

Abstract

Why be moral? One possible, and compelling answer is that to act morally is in an agent's self-interest. Such an answer can be either elevationist (broadly speaking the Aristotelian/Platonic approach) where self-interest is elevated to coincide with living the good life, or reductionist where morality is defined as acting in an agent's self-interest.

Elevationist moral theories appear flawed. If you are in possession of information that, if divulged, will bring about the deaths of others then it may be virtuous to stay silent. However, if staying silent results in you being slowly tortured to death in an effort to extract the information then it seems bizarre to suggest that in doing so you are flourishing, happy, or acting out of self-interest.

Reductionist moral theories, acting for the 'good of self' rather than the 'good of others', are widely considered to be the antithesis of morality. Moral philosophers tend to attack such positions claiming that the doctrine of egoism is unworkable. It is commonly claimed that any theory which recommends 'an agent do x if x is in the agent's best interest' is inconsistent, incoherent, or contradictory and fails to meet the basic requirements of a moral theory (notably the requirement of universalisability).

I begin this thesis with an examination of ethical egoism in its most widely known consequentialist form; i.e. an agent ought to act so as to bring about the best consequences for that agent. I examine the major criticisms of this theory and demonstrate that the axioms of egoism can be developed so as to overcome these criticisms. I argue that consequentialist based ethical egoism is coherent, consistent and noncontradictory. However, I go on to argue that while egoism can be formulated in a manner that overcomes all the aforementioned analytic criticisms it is a flawed moral theory in that within certain contexts the action deemed morally correct by egoism is, as a matter of fact, morally pernicious. That a theory contains a flaw is not reason enough to discard the entire theory and I go on to contend that the problem with egoism is the consequentialist approach, not the fact that it is based on self-interest.

In Part 2 of the thesis I abandon the consequentialist approach and examine the possibility of a flourishing-based form of ethical egoism. I further develop the axioms of egoism established in Part 1 through an examination of the concept of flourishing (as commonly associated with virtue ethics). Ultimately I tread a path between the consequentialist and elevationist positions. While I do not elevate self-interest to acting virtuously I do contend that an egoist must adopt certain virtues if that egoist is to have the best possibility to flourish. However, I further contend that an egoist ought to act so as to promote that which the egoist values and that this agent-relative hierarchy of values, which necessarily contains certain virtues, determines the manner in which an egoist ought to act.

*For Deb
and out of Self-Interest*

Table of Contents

Abstract	ii
Part 1	
Consequentialist Egoism	
Chapter 1 Egoism	
§1.0 Introduction	2
§1.1 Egoism – A Case Study	5
§1.2 Egoism – Criticisms	16
§1.2.1 Egoism is Self Defeating	16
§1.2.2 Definitional Issues	18
§1.2.3.1 Logical (Internal) Contradictions	20
§1.2.3.2 Logical (External) Contradictions	21
§1.2.4 Inconsistency / Incoherence	21
§1.2.5 The Parity of Morality and the Nature of the Self	23
§1.2.6 Defending Egoism	25
§1.2.7 Egoism is a Pernicious Doctrine	26
Chapter 2 Definitional Issues	
§2.0 Introduction	27
§2.1 Peter Singer’s Utilitarianism	28
§2.2 Rights and Virtues	29
§2.3 Appeals to the Best Explanation	29
§2.4 Defining Morality; Self vs Others	34
§2.4.1 The Lack of Self Guidance	35
§2.4.2 A Question of (Moral) Effort	36
§2.4.3 The Myth of Pure Motivation	38
§2.4.4 Inconsistencies in Defining Morality as Essentially Other-Regarding	39
§2.4.5 Admiration of Self-Interest	41
§2.5 Conflict Resolution	44
§2.6 Conclusion	50

Chapter 3 Egoism is Contradictory	
§3.0 Ultimate Rational Ends and the Universal Good	52
§3.1 Denying the Importance of Logical Contradiction	57
§3.2 Broad's Response to Moore	57
§3.3 Smith's Response to Moore	59
§3.4 Rasmussen and Relative Values	63
§3.5 Moral Relativism	65
§3.6 Moorean Challenges and Naturalistic Fallacies	68
§3.7 Conclusion	70
Chapter 4 Inconsistency and Incoherence	
§4.0 Introduction	72
§4.1 Ultimate Principles and Ethical Egoism	72
§4.1.1 Medlin's Attack on Ethical Egoism	74
§4.1.2 Inconsistency and Incoherence	76
§4.2 Kalin in Defence of Ethical Egoism	78
§4.2.1 Kalin, Wittgenstein, and Private Morality	80
§4.2.2 Kalin, Carlson, and What the Egoist Wants	82
§4.2.3 Private Universal Egoism	85
§4.3 Re-universalising Ethical Egoism	85
§4.4 Ethical Egoism and Moral Motivation	88
§4.4.1 Internal and External Reasons	89
§4.5 Ethical Egoism and the Sports Model	91
§4.6 Conclusion	95
Chapter 5 Further Inconsistencies and Incoherence	
§5.0 Introduction	97
§5.1 The Prisoners' Dilemma	99
§5.1.1 Iterated Games Theory	103
§5.1.2 The Prisoners' Dilemma, the Egoist, and Tit For Tat	105
§5.1.3 A Final Word on Game Theory	107
§5.1.3.1 Gauthier – The Impossibility of Rational Egoism	108

§5.2 The Nature of Self	113
§5.2.1 Zemach's Self	113
§5.2.1.1 Egoism Without the Self	115
§5.2.1.2 Egoism, the Self, and Worldviews	116
§5.2.2 The Self in Time	118
§5.2.3 The Presentist Challenge	120
§5.2.4 Conclusion	122
Chapter 6 Moral Facts	
§6.0 Is Egoism a Pernicious Doctrine?	123
§6.1 Rachel's Argument	124
§6.2 Harman's Thesis	125
§6.3 Appeals to the Best Explanation and Evaluative Facts	127
§6.4 The Vapour Trail and the Best Explanation	135
§6.5 Moral Facts and Best Explanations	136
§6.6 Counterfactuals	139
§6.7 Artificially Constructed Asymmetry	140
§6.8 Mathematics	143
§6.9 Best Explanation – One Last Time	146
§6.10 Conclusion	150
Chapter 7 Consequentialism, Egoism, and Pernicious Doctrines	
§7.0 Introduction	152
§7.1 Question Begging	152
§7.1.1 Rachels' Defence	153
§7.1.2 Question Begging at a Deeper Level	155
§7.1.3 The Power of Intuition	156
§7.1.3.1 Temporally Relative Moral Intuitions	157
§7.1.3.2 Moral Intuitions and Certainty	158
§7.1.4 Moral Facts	159
§7.2 Defining Terms	161
§7.2.1 The Egoist's 'Interests'	162
§7.2.2 'To Best Promote'	163

§7.2.3 The Extended Argument	164
§7.3 Machan’s Objection	165
§7.4 Hypothetical Cases	168
§7.4.1 Assessing Hypotheticals – Egoism	168
§7.4.1.1 Denying Temporal Importance	169
§7.4.1.2 Accepting All Temporal Assessments as Important	169
§7.4.1.3 Defining the Moment of Importance	170
§7.4.2 Assessing Hypotheticals – Generalization	171
§7.5 Temporality Again – Actual and Probable Consequences	173
§7.6 Calculating Probable Consequences	174
§7.6.1 Take the Money and Run	175
§7.6.2 Egoistic Benevolence	176
§7.6.3 Assessing the Options	177
§7.7 The Final Formulation of Rachels’ Argument	177

Part 2

Nonconsequentialist Egoism

Chapter 8 Flourishing

§8.0 Introduction	181
§8.1 The Concept of Flourishing	184
§8.1.1 Flourishing and the Complexity of Life	186
§8.2 Flourishing – A Two-Factor Account	193
§8.2.1 The First Factor – Eco Flourishing	194
§8.2.2 The Second Factor – ‘My’ Flourishing	197

Chapter 9 Rand’s Ethics

§9.0 Introduction	200
§9.1 On the Randian Argument	204
§9.1.1 Values	204
§9.1.2 Life as a Value	208
§9.1.3 Rational Persons	212
§9.1.3.1 A Brief Outline of the Virtues	214

§9.1.4 Social Rights	215
§9.1.5 Happiness and the Good of Others	218
§9.2 Constrained Egoism	220
§9.3 A Quick Criticism of Randian/Flourishing Egoism	222
§9.4 Conclusion	223
Chapter 10 Flourishing Egoism	
§10.0 Introduction	224
§10.1 Beings of a Certain Kind and Universal Values	225
§10.1.1 An Agent's Worldview and Holistic Values	227
§10.1.2 The Importance of Honesty	232
§10.1.3 The Importance of Respect for Others	235
§10.1.4 The Importance of Limited Benevolence	236
§10.1.5 The Requirement of Temperance	238
§10.2 Returning to Rand – Formalizing the Principles	239
§10.3 Subjective Values	241
§10.4 Initial Objections	244
§10.4.1 Ethical Egoism in Action: The Lifeboat Dilemma	246
§10.4.1 Ethical Egoism in Action: Rachels' Doctor	247
Chapter 11 A Day in Punxsutawney	
§11.0 Introduction	249
§11.1 A Day in Punxsutawney: The Scenario	249
§11.2 Denial and Anger	251
§11.3 Bargaining	251
§11.4 Depression	252
§11.5 Acceptance	253
§11.5.1 Groundhog Day: The Ballad of the Tramp	254
§11.5.2 Groundhog Day: The Boy and the Tree	256
§11.5.3 Groundhog Day: The Value of Music	256
§11.6 Phil and the Values of Flourishing-Egoism	259
§11.7 Conclusion	261
Final Conclusion	
Out of Self-Interest	263

VIRTUE AND SELF-INTEREST

A thesis
submitted in fulfilment
of the requirements for the degree
of
Doctor of Philosophy in Philosophy
at
The University of Waikato
by
TERY HARDWICKE

Abstract

Why be moral? One possible, and compelling answer is that to act morally is in an agent's self-interest. Such an answer can be either elevationist (broadly speaking the Aristotelian/Platonic approach) where self-interest is elevated to coincide with living the good life, or reductionist where morality is defined as acting in an agent's self-interest.

Elevationist moral theories appear flawed. If you are in possession of information that, if divulged, will bring about the deaths of others then it may be virtuous to stay silent. However, if staying silent results in you being slowly tortured to death in an effort to extract the information then it seems bizarre to suggest that in doing so you are flourishing, happy, or acting out of self-interest.

Reductionist moral theories, acting for the 'good of self' rather than the 'good of others', are widely considered to be the antithesis of morality. Moral philosophers tend to attack such positions claiming that the doctrine of egoism is unworkable. It is commonly claimed that any theory which recommends 'an agent do x if x is in the agent's best interest' is inconsistent, incoherent, or contradictory and fails to meet the basic requirements of a moral theory (notably the requirement of universalisability).

I begin this thesis with an examination of ethical egoism in its most widely known consequentialist form; i.e. an agent ought to act so as to bring about the best consequences for that agent. I examine the major criticisms of this theory and demonstrate that the axioms of egoism can be developed so as to overcome these criticisms. I argue that consequentialist based ethical egoism is coherent, consistent and noncontradictory. However, I go on to argue that while egoism can be formulated in a manner that overcomes all the aforementioned analytic criticisms it is a flawed moral theory in that within certain contexts the action deemed morally correct by egoism is, as a matter of fact, morally pernicious. That a theory contains a flaw is not reason enough to discard the entire theory and I go on to contend that the problem with egoism is the consequentialist approach, not the fact that it is based on self-interest.

In Part 2 of the thesis I abandon the consequentialist approach and examine the possibility of a flourishing-based form of ethical egoism. I further develop the axioms of egoism established in Part 1 through an examination of the concept of flourishing (as commonly associated with virtue ethics). Ultimately I tread a path between the consequentialist and elevationist positions. While I do not elevate self-interest to acting virtuously I do contend that an egoist must adopt certain virtues if that egoist is to have the best possibility to flourish. However, I further contend that an egoist ought to act so as to promote that which the egoist values and that this agent-relative hierarchy of values, which necessarily contains certain virtues, determines the manner in which an egoist ought to act.

*For Deb
and out of Self-Interest*

Table of Contents

Abstract	ii
Part 1	
Consequentialist Egoism	
Chapter 1 Egoism	
§1.0 Introduction	2
§1.1 Egoism – A Case Study	5
§1.2 Egoism – Criticisms	16
§1.2.1 Egoism is Self Defeating	16
§1.2.2 Definitional Issues	18
§1.2.3.1 Logical (Internal) Contradictions	20
§1.2.3.2 Logical (External) Contradictions	21
§1.2.4 Inconsistency / Incoherence	21
§1.2.5 The Parity of Morality and the Nature of the Self	23
§1.2.6 Defending Egoism	25
§1.2.7 Egoism is a Pernicious Doctrine	26
Chapter 2 Definitional Issues	
§2.0 Introduction	27
§2.1 Peter Singer’s Utilitarianism	28
§2.2 Rights and Virtues	29
§2.3 Appeals to the Best Explanation	29
§2.4 Defining Morality; Self vs Others	34
§2.4.1 The Lack of Self Guidance	35
§2.4.2 A Question of (Moral) Effort	36
§2.4.3 The Myth of Pure Motivation	38
§2.4.4 Inconsistencies in Defining Morality as Essentially Other-Regarding	39
§2.4.5 Admiration of Self-Interest	41
§2.5 Conflict Resolution	44
§2.6 Conclusion	50

Chapter 3 Egoism is Contradictory	
§3.0 Ultimate Rational Ends and the Universal Good	52
§3.1 Denying the Importance of Logical Contradiction	57
§3.2 Broad's Response to Moore	57
§3.3 Smith's Response to Moore	59
§3.4 Rasmussen and Relative Values	63
§3.5 Moral Relativism	65
§3.6 Moorean Challenges and Naturalistic Fallacies	68
§3.7 Conclusion	70
Chapter 4 Inconsistency and Incoherence	
§4.0 Introduction	72
§4.1 Ultimate Principles and Ethical Egoism	72
§4.1.1 Medlin's Attack on Ethical Egoism	74
§4.1.2 Inconsistency and Incoherence	76
§4.2 Kalin in Defence of Ethical Egoism	78
§4.2.1 Kalin, Wittgenstein, and Private Morality	80
§4.2.2 Kalin, Carlson, and What the Egoist Wants	82
§4.2.3 Private Universal Egoism	85
§4.3 Re-universalising Ethical Egoism	85
§4.4 Ethical Egoism and Moral Motivation	88
§4.4.1 Internal and External Reasons	89
§4.5 Ethical Egoism and the Sports Model	91
§4.6 Conclusion	95
Chapter 5 Further Inconsistencies and Incoherence	
§5.0 Introduction	97
§5.1 The Prisoners' Dilemma	99
§5.1.1 Iterated Games Theory	103
§5.1.2 The Prisoners' Dilemma, the Egoist, and Tit For Tat	105
§5.1.3 A Final Word on Game Theory	107
§5.1.3.1 Gauthier – The Impossibility of Rational Egoism	108

§5.2 The Nature of Self	113
§5.2.1 Zemach's Self	113
§5.2.1.1 Egoism Without the Self	115
§5.2.1.2 Egoism, the Self, and Worldviews	116
§5.2.2 The Self in Time	118
§5.2.3 The Presentist Challenge	120
§5.2.4 Conclusion	122
Chapter 6 Moral Facts	
§6.0 Is Egoism a Pernicious Doctrine?	123
§6.1 Rachel's Argument	124
§6.2 Harman's Thesis	125
§6.3 Appeals to the Best Explanation and Evaluative Facts	127
§6.4 The Vapour Trail and the Best Explanation	135
§6.5 Moral Facts and Best Explanations	136
§6.6 Counterfactuals	139
§6.7 Artificially Constructed Asymmetry	140
§6.8 Mathematics	143
§6.9 Best Explanation – One Last Time	146
§6.10 Conclusion	150
Chapter 7 Consequentialism, Egoism, and Pernicious Doctrines	
§7.0 Introduction	152
§7.1 Question Begging	152
§7.1.1 Rachels' Defence	153
§7.1.2 Question Begging at a Deeper Level	155
§7.1.3 The Power of Intuition	156
§7.1.3.1 Temporally Relative Moral Intuitions	157
§7.1.3.2 Moral Intuitions and Certainty	158
§7.1.4 Moral Facts	159
§7.2 Defining Terms	161
§7.2.1 The Egoist's 'Interests'	162
§7.2.2 'To Best Promote'	163

§7.2.3 The Extended Argument	164
§7.3 Machan’s Objection	165
§7.4 Hypothetical Cases	168
§7.4.1 Assessing Hypotheticals – Egoism	168
§7.4.1.1 Denying Temporal Importance	169
§7.4.1.2 Accepting All Temporal Assessments as Important	169
§7.4.1.3 Defining the Moment of Importance	170
§7.4.2 Assessing Hypotheticals – Generalization	171
§7.5 Temporality Again – Actual and Probable Consequences	173
§7.6 Calculating Probable Consequences	174
§7.6.1 Take the Money and Run	175
§7.6.2 Egoistic Benevolence	176
§7.6.3 Assessing the Options	177
§7.7 The Final Formulation of Rachels’ Argument	177

Part 2

Nonconsequentialist Egoism

Chapter 8 Flourishing

§8.0 Introduction	181
§8.1 The Concept of Flourishing	184
§8.1.1 Flourishing and the Complexity of Life	186
§8.2 Flourishing – A Two-Factor Account	193
§8.2.1 The First Factor – Eco Flourishing	194
§8.2.2 The Second Factor – ‘My’ Flourishing	197

Chapter 9 Rand’s Ethics

§9.0 Introduction	200
§9.1 On the Randian Argument	204
§9.1.1 Values	204
§9.1.2 Life as a Value	208
§9.1.3 Rational Persons	212
§9.1.3.1 A Brief Outline of the Virtues	214

§9.1.4 Social Rights	215
§9.1.5 Happiness and the Good of Others	218
§9.2 Constrained Egoism	220
§9.3 A Quick Criticism of Randian/Flourishing Egoism	222
§9.4 Conclusion	223
Chapter 10 Flourishing Egoism	
§10.0 Introduction	224
§10.1 Beings of a Certain Kind and Universal Values	225
§10.1.1 An Agent's Worldview and Holistic Values	227
§10.1.2 The Importance of Honesty	232
§10.1.3 The Importance of Respect for Others	235
§10.1.4 The Importance of Limited Benevolence	236
§10.1.5 The Requirement of Temperance	238
§10.2 Returning to Rand – Formalizing the Principles	239
§10.3 Subjective Values	241
§10.4 Initial Objections	244
§10.4.1 Ethical Egoism in Action: The Lifeboat Dilemma	246
§10.4.1 Ethical Egoism in Action: Rachels' Doctor	247
Chapter 11 A Day in Punxsutawney	
§11.0 Introduction	249
§11.1 A Day in Punxsutawney: The Scenario	249
§11.2 Denial and Anger	251
§11.3 Bargaining	251
§11.4 Depression	252
§11.5 Acceptance	253
§11.5.1 Groundhog Day: The Ballad of the Tramp	254
§11.5.2 Groundhog Day: The Boy and the Tree	256
§11.5.3 Groundhog Day: The Value of Music	256
§11.6 Phil and the Values of Flourishing-Egoism	259
§11.7 Conclusion	261
Final Conclusion	
Out of Self-Interest	263

VIRTUE AND SELF-INTEREST

Part 1

Consequentialist Egoism

Chapter 1 Egoism

§1.0 Introduction

Egoism can be divided into two general types of theories: psychological egoism and ethical (or rational) egoism. Psychological egoism is a descriptive theory claiming that we always, as a matter of fact, act in the manner we believe will further our own self-interest. Any act that appears altruistic in nature can always be explained as having been motivated by the agent's self-interest. A psychological egoist may claim that performing act 'X' made the agent happier and that the agent's real motivation for 'X', which on the surface seems to be altruistic considerations, was the agent's own happiness and not the good of another. The example of Abraham Lincoln and the drowning pigs demonstrates such a position;

Mr. Lincoln once remarked to a fellow passenger on an old-time mud coach that all men were prompted by selfishness in doing good. His fellow passenger was antagonizing this position when they were passing over a corduroy bridge that spanned a slough. As they crossed this bridge they espied an old razor-backed sow on the bank making a terrible noise because her pigs had got into the slough and were in danger of drowning. As the old coach began to climb the hill, Mr. Lincoln called out, "Driver, can't you stop a moment?" Then Mr. Lincoln jumped out, ran back and lifted the little pigs out of the mud and water and placed them on the bank. When he returned, his companion remarked, "Now Abe, where does selfishness come in on this little episode?" "Why bless your soul, Ed, that was the very essence of selfishness. I should have had no peace of mind all day had I gone and left that suffering old sow worrying over those pigs. I did it to get peace of mind, don't you see?" (Quoted in J. Feinberg, "Reason and Responsibility")

Lincoln claims he was acting to promote his own happiness by the removal of a situation that caused him stress. An apparently altruistic act of kindness is therefore claimed to have been motivated by selfishness. However, in many cases psychological egoism – i.e. the claim that all human actions are motivated by selfish desires - is implausible. Shaver (2002) tells the story of a soldier who throws himself on a grenade to prevent others in his platoon being killed; it seems absurd to suggest that such an act could be described as motivated by the soldier's self-interest. The psychological egoist might offer one of two possible explanations for the soldier's actions. Firstly, it might be claimed the soldier's

action was from self-interest in that the soldier could not have lived with himself had he acted otherwise. What I presume is meant by ‘not being able to live with himself’ is that the soldier would feel guilt and remorse at not having acted to save the lives of the others. As Shaver (2002, para. 7) notes, “guilt may presuppose that the soldier had a non-self-regarding desire for doing what he takes to be right”. The soldier’s guilt avoidance appears not to be out of purely selfish motives but from a feeling of not living up to his obligations to others. Perhaps a similar criticism might be laid against Lincoln’s analysis of why he was motivated to act.

Secondly, the psychological egoist might claim that given that the soldier did jump on the grenade this action must have been motivated by what the soldier believed was in his best interest. If this is what is meant by psychological egoism then the position is unfalsifiable, for it is simply the claim that:

1. Any and every action we take is chosen by us
2. *therefore* in acting as we chose we must have been doing what we believed was in our self-interest
3. *therefore* we acted from our own self-interest.

“[P]sychological egoism turns out to be trivially true” if, by definition, any and all “intentional action is self-interested” (Shaver, 2002). Psychological egoism offers a reductionist account of self-interest; self-interest is simply “whatever one is finally, all things considered, motivated to do” (Baier, 1991, p. 201). According to psychological egoism the question, ‘What ought I do in this context?’ is meaningless (logically, it can have no objectively correct answer). Any conclusion the agent reaches in answering this question, where the answer is denoted by the behavior of the agent, is necessarily ‘what the agent believed it was in the agent’s self-interest to do’. Whatever action the agent undertakes (assuming the agent was free to choose) will, as a matter of fact, be what the agent believed was in the agent’s best interest. Yet, agents do ponder this very question for it appears, intuitively, to be very meaningful indeed. Given that moral agents do seek answers to this question it seems reasonable to search for better ways of answering it. Saying ‘A did X therefore X was, as a matter of fact, what A

believed was in A's best interest' is irrelevant to how we choose, and sometimes agonise over, our actions.

Ethical (or rational) egoism is a theory of how we ought to act, as opposed to how we do act, and refers to any normative theory that:

- (i) Tells us that we ought to act, or that we have sufficient reason to act, in a certain manner, and
- (ii) Preferences the self, self-interest, self-happiness or self-value over others or other concerns, in determining how we ought to act.

Ethical egoism, free from the rigours of philosophical analysis, seems an entirely plausible theory¹. Doing something because it is in our best interests to do so appears intuitively right whereas an altruistic action, especially when doing so goes against self-interest, calls for justification. Butler (1726, Sermon II, Para. 20) noted that “when we sit down in a cool hour, we can neither justify to ourselves this or any other pursuit till we are convinced that it will be for our happiness, or at least not contrary to it”.

However, that a theory appeals to a layperson is far from sufficient grounds for members of the philosophical community to adopt it. Most contemporary moral philosophers view the egoist standpoint as reprehensible and the antithesis of morality, an antithesis that must be defeated if moral theory is to make any progress². Given this general presupposition that egoism *is* false, those criticizing the theory tend to spend very little time developing its premises. The result is that a theory of selfishness that few (if anyone) would actually subscribe to is shown to be false. For example, according to Rachels (1974, p. 297) “[by] ethical egoism, each of us should take the attitude that other people simply don't matter, except insofar as they are useful to us” and, “the right thing to do, on any occasion, is whatever would best promote [the egoist's] best interests” (Rachels, 1978, p. 426). Dismissing such a theory shows only that selfishness, using others solely as a means to our own ends, and self-interest with total disregard for others,

¹ I will draw no distinction between ethical and rational egoism at this early stage.

² In as much as, if the question ‘Why be moral?’ can't be answered there seems little point in working out the intricacies of how to be moral.

is flawed as the basis of a moral theory. However, it is far from clear that this is what is entailed by an ethical egoist standpoint.

In determining exactly what is entailed by ethical egoism I will briefly digress and consider just how an egoist might act when facing a moral dilemma. My reason for doing so is to show that while the egoist acts from self-interested motivation it does not necessarily follow that the egoist's actions will be totally selfish in nature or that the actions will be far different from those recommended by other moral theories³. Following the case study I will introduce the central criticisms of egoism and in discussing these criticisms will determine what the axioms of a valid, noncontradictory, and consistent theory of ethical egoism might be.

§1.1 Egoism – A Case Study

Twenty passengers are out on a pleasure cruise when the boat runs into serious trouble and starts to sink. Unfortunately, one of the two lifeboats is in disrepair and unusable. One female passenger (the egoist) manages to successfully launch the usable lifeboat and discovers that all of the other passengers have jumped overboard and are now swimming around calling out to her for help. The cruise is not expected to return for 72 hours and so the passengers of the boat will not be missed for three days and none of the 19 passengers, now in the water, will be able to survive this period unless taken aboard the lifeboat. The lifeboat is well equipped and has provisions that will guarantee the safety of five passengers until help arrives. Every passenger above the five (the absolute safety margin) who are taken on board reduces the overall chance of survival by 10%; i.e. eight passengers in the lifeboat will have a 70% chance of surviving until rescued, 12 passengers will have a 30% chance and, if 15 or more passengers are taken aboard then it is certain that all will perish.

Now, assuming she has the means to stop others boarding the lifeboat without her permission, how many others (if any) ought the egoist to allow on board?

Critics of egoism often assume that if the egoist is looking out for her own best interests then she will want to do the utmost to protect her own life. The egoist would simply sail off into the sunset leaving 19 passengers to die or, at best, might rescue four other like-minded souls if they were needed to help man the

³ To be selfish is to act for personal gain; to act out of self-interest may require the foregoing of personal gain. I will examine the distinction between acting from selfishness and acting out of self-interest in section 2.

boat. (They would need to be like-minded otherwise it might prove very difficult to convince the other four that they should put themselves at no risk and rescue no others – she may very quickly find that she has lost control of the situation, no longer commands the life raft, and is unable to effectively act to maximise her interests.)

However, the egoist's decision is not so simple; she will also want to ensure that upon rescue she is not held criminally responsible for allowing nineteen to die - a result that is likely to be highly detrimental to her interests and will result in harsh repercussions. (In order to avoid making the calculations unnecessarily complex I will assume that the repercussions are 'harsh' to the extent that they are equal in value to 'loss of life'). Therefore, she should certainly rescue at least four others, as this entails no risk to herself. However, even assuming the difficult task of surveying the nineteen for four like-minded persons could be accomplished, it is still probable that she would be judged as acting wrongly and punished harshly upon rescue.

It might be suggested at this point that if egoism is correct then the laws (upon which the harsh punishments are based) require that an agent acts immorally (i.e. against the agent's self-interest). Such an assumption is somewhat premature; until the axioms of ethical egoism have been fully expounded it cannot be pre-determined that laws based on egoistic principles would be so different from laws based on other principles.

Returning to the question of who the egoist ought to rescue I suggest that, for the egoist, the answer lies in balancing probabilities. That is, balancing the probable outcome of specific actions and calculating the degree to which they are beneficial or detrimental to the egoist. On the one hand we have the probability of the egoist's survival; on the other the probability of the egoist facing repercussions upon rescue (and the probability that those taken on board rebel against the egoist's decision). It is difficult to assign probability values; however, I have made the following assumptions:

1. The totally selfish action (the egoist ensuring she has a 100% chance of survival) is 100% certain to result in severe repercussions.
2. Maximum risk (rescuing 14) ensures no repercussions.
3. The median lies around the 50% mark and few would judge someone harshly who accepted a 50/50 chance of survival.

It is of course likely that the degree of repercussion would also vary and that in some cases repercussions would be acceptable to the egoist given that it is better to be alive, even if suffering hardship, than it is to be dead⁴. However, accounting for this would add an extra degree of complexity to the calculations. Further, the extra consideration is unnecessary for examining the point at hand. So I will add the premise:

4. The repercussions, at all levels, are severe to the point that the egoist will want to do her utmost to avoid them. That is, the repercussions are equal in value to loss of life.

Using these variables the egoist's decision can be reached mathematically;

$$S\% - (S\% \cdot R\%)$$

S% is the probability of survival and R% is the probability of repercussion.

This formula calculates the probability of survival (S%) less (S% times R%), the chance of survival, multiplied by the probability of repercussions. Repercussions are multiplied by survival probability for the simple reason that if the egoist does not survive then the egoist faces no probability of repercussions (assuming that repercussions against the dead, or the estate of the dead, are not of any interest to the egoist)⁵.

Calculating the probability that the egoist will face repercussions is extremely difficult and any claims that the 'actual probabilities are X%' will be arbitrary. However, as in any consequentialist theory, if a calculation is to be made then values must be assigned. Therefore, I have selected three mathematical methods of distribution, which I believe would be generally accepted, to apportion the

⁴ This point is somewhat controversial as some may claim that death is better than a life of extreme suffering. However, it might be countered that a life that has even the smallest possibility of value outranks death and the certainty of no value. All I wish to claim here is that given everyday circumstances life is preferable to death even if that life involves a degree of suffering.

⁵ I have taken a fairly general game theory approach to the calculations, according to which the value to an agent is the chance of something good (survival) less the chance of something bad (repercussions). The same results are obtained by the more traditional mathematical approach of multiplying positive probabilities, i.e. the chance of survival multiplied by the chance of avoiding repercussions. $S\% \cdot (1-R\%)$

probability of repercussions between 0% (for taking maximal risk) and 100% (for taking no risk).

Fig 1.1: Acting so as to ensure a chance of survival (S%) determines the probabilities of repercussion (R%)

Saved	Chances of survival	Probability of repercussions (R%) ⁶		
		Curve	Reduced C	Incremental
1-5	100%	100%	100%	100%
6	90%	65%	71%	80%
7	80%	36%	45%	60%
8	70%	16%	23%	40%
9	60%	4%	6%	20%
10	50%	No Repercussions		
11	40%	No Repercussions		
12	30%	No Repercussions		
13	20%	No Repercussions		
14	10%	No Repercussions		
15	0% (Death)	No Repercussions		

I have assumed in the above figures that no repercussions occur when a 50% risk is accepted. However, setting a 0% probability of 'no repercussions' only when the maximum risk (save 14 passengers at a risk of 90%) is accepted results in only minor differences to the final result. The Curve method of distribution assumes that the chance of repercussion drops quickly as risk is accepted. Incremental assumes the probability of repercussion is directly proportional to the degree of risk taken and the reduced curve (Reduced C) lies between the two.

If these values are correct then the result of the equation $[S\% - (S\% \cdot R\%)]$ will be that if the egoist rescues seven passengers (including herself), thus reducing her chance of survival to 80%, she accepts a 36% to 60% probability of repercussions and the value of her action will be between 0.512 $[80\% - (80\% \cdot 36\%)]$ and 0.32 $[80\% - (80\% \cdot 60\%)]$.

⁶ The three methods of assigning value to repercussions are Curve (even ratio squared), Reduced Curve (non-even ratio squared), and Incremental (Values are staggered at an even ratio between 0% and 100%).

Assessing how many people the sole passenger of the lifeboat ought to save is now simply a matter of performing this calculation ten times. Saving one to five passengers at no risk produces the same result and there is no point doing the calculation for 15 or more passengers as this results in certain death for all.

Fig 1.2: The value to the egoist of saving n passengers.

The Value of the action:			
Egoist saves n passengers	Curve	Reduced C	Incremental
One-five	0	0	0
Six	0.324	0.26	0.18
Seven	0.512	0.44	0.32
Eight	0.588 *	0.54	0.42
Nine	0.576	0.56 *	0.48
Ten	0.5	0.5	0.5 *
Eleven	0.384	0.36	0.4
Twelve	0.252	0.14	0.3
Thirteen	0.128	Negative	0.2
Fourteen	0.036	Negative	0.1
Fifteen	All Die	All Die	All Die

* Indicates the optimal value under each system of calculation.

The egoist then ought to save between eight and ten passengers and it maximises her self-interest to subject herself to a risk of between 30% and 50%. It may be insisted that the only way the egoist can be certain of avoiding repercussions is to rescue 14 passengers. However, this makes very little difference to the final result as is demonstrated in the following tables.

Fig 1.3: Acting so as to ensure a chance of survival (S%) determines the probabilities of repercussion (R%) (Fig 1.1 revised)

Saved	Chances of survival	Probability of repercussions (R%)	
		Curve	Incremental
1-5	100%	100%	100%
6	90%	79%	89%
7	80%	60%	78%
8	70%	44%	67%
9	60%	31%	55%
10	50%	20%	44%
11	40%	11%	33%
12	30%	5%	22%
13	20%	1%	11%
14	10%	No Repercussions	
15	All Die	No Repercussions	

Fig 1.4: The value to the egoist of saving n passengers. (Revised)

Saves	The Value Of the action	
	Curve	Incremental
One-five	0.000	0.00
Eight	0.388	0.233
Nine	0.415	0.266
Ten	0.401	0.278
Eleven	0.355	0.266
Fourteen	0.100	0.10

Revising the figures (to allow for the possibility of repercussion if any fewer than the maximum number possible are rescued) means that under the curve-method of distribution the egoist is obligated to save nine (rather than eight) passengers and under the incremental method of assigning probabilities the egoist is obligated to save ten passengers (as per the previous calculation).

Some may object to the means by which the egoist has reached this decision, and I do not intend to debate that particular issue at this time. My objective was simply

to show that the egoist's decision differs from what many assume it will be (selfish, 100% self preservation) and that the action promoted by the egoist's doctrine, divorced from the reasons for the action, may not differ all that much from the correct action as determined by other moral theories and appears intuitively to be a morally reasonable course of action.

I am not claiming there are no differences between the moral theories. While, according to egoism, selfishly saving only oneself is not *the* correct act (i.e. it is not the act that maximises the agent's self-interest), it does hold the view that there is no moral difference between (a) saving only herself and (b) saving four others in so far as the acts are assigned the same utility. Such a position conflicts with what I would assume normal moral intuition to be; namely, saving more is better than saving fewer. However, it remains to be seen whether it matters that egoism draws no distinction in the degree of wrongness of the two acts, given that the act of saving only herself and the act of saving four others are equally judged as wrong⁷. For now I will turn my attention to considering what action might be recommended by alternative moral theories.

The agent subscribing to virtue theory ought to find some mean between acting cowardly (saving few if any passengers) and acting with foolhardiness⁸. The ideally virtuous person would attempt to find a mean between these extremities and act courageously. It is difficult to determine exactly what the courageous act would be, however if the courageous action is the mean between cowardliness and rashness we can make a rough calculation. If we assume the cowardly act to be saving five or fewer and the foolhardy act to be an attempt to save anything above 14 passengers then we can assume the mean (and the courageous act) to be attempting to save ten passengers⁹. The virtuous agent would probably also act to reduce the suffering of those left behind but then so might the egoist, albeit for very different reasons.

⁷ From this point on the term 'egoism' refers to ethical egoism.

⁸ A foolhardy act would be attempting to save all of the passengers such that the outcome is one of complete catastrophe.

⁹ Coward (5) + Foolhardy (15) divided by 2.

The utilitarian has as difficult a calculation as the egoist - how many to attempt to save in an effort to maximise interests. Maximising overall utility requires maximising positive utility (the probability of survival of those taken into the lifeboat) and minimising negative utility (the number of unfortunates left behind and denied the possibility of survival). This can be calculated by:

$$\text{Utility} = S\%n - \Sigma(S\%_x d_x)$$

Where $S\%n$ represents n persons aboard the lifeboat with $S\%$ possibility of survival and $\Sigma(S\%_x d_x)$ represents the negative utility incurred from each passenger (d_x), left to perish and denied $S\%_x$ probability of survival. The life raft can support a maximum of 14 passengers therefore the maximum value of d_x is 13¹⁰.

This builds risk aversion into the utilitarian calculation, in so far as the utility value of attempting to rescue n passengers is reduced according to the probability of success. To save four passengers (five including the utilitarian) with a 100% chance of survival produces a positive utility of 5.0, however the next person who could have been saved but was left to perish produces a negative utility of 0.9, the next 0.8, 0.7, 0.6 etc. This results in a total negative utility of 4.5. Thus the total utility value of saving five of the twenty passengers is 0.5 (5.0 positive utility less 4.5 negative utility).

Saving a total of nine passengers and leaving 11 to die (the six the life boat cannot support plus the five who could have been rescued) means that those on board the lifeboat have a 60% probability of survival. The utility of this act can be calculated by $(5.4) - (0.5+0.4+0.3+0.2+0.1)$, The 60% chance the nine have of survival gives a positive utility of 5.4, the five who could have been saved but were not produces a negative utility of 1.5, and the overall utility value of saving nine passengers is 3.9. As we did with the egoist we can now calculate the value for all of the possible outcomes:

¹⁰ The 20 passengers of the cruise, less the six that cannot be saved, less the person already occupying the lifeboat.

Fig 1.5 Utilitarian value of saving n passengers

Passengers Saved	Utilitarian Value
5	0.5
6	1.8
7	2.8
8	3.5
9	3.9
10	4.0
11	3.8
12	3.3
13	2.5
14	1.4
15	0.0

The result of such an equation obligates the utilitarian to an action similar to the one reached by the egoist (save nine or ten) and the same as the virtuous person (save ten). According to the utilitarian one ought to save ten passengers and accept odds of survival of 50%. In this scenario, maximising self-interest and maximising the interests of the majority result in very similar actions. Without substantial additional assumptions about how conflicting rights are to be assessed there is no equivalent calculation that can be made for a rights based approach.

Somewhat surprisingly, it seems that if we consider who the egoist and utilitarian would save we find that the egoist's, rather than the utilitarian's, action matches more closely with what intuitively seems to be the correct thing to do. Most people may find the utilitarian's motives, in working for the majority, intuitively more plausible. However, our egoist would be likely to rescue some women and children (assuming some were onboard) and while her reasons for doing so might draw scorn, (that rescuing those most vulnerable is likely to further reduce the chance of repercussions), the act itself, rescuing *women and children first*, fits more closely with commonly held beliefs about the right thing to do. On the other hand, while acting so as to give the greatest number the chance of survival, the utilitarian ought to select those more able to staff the lifeboat - the fitter and

stronger. Small children, and many women, are unlikely to qualify; in fact the utilitarian may well find herself in a position where she is compelled to throw herself overboard to maximise chances of survival for the group.

My purpose in discussing this case study is to show, in the first place, that an egoist theory *may*, at least in some contexts, promote actions that would gain wide approval (and it must be emphasised I am referring only to the act itself not the motivation for the act). Further, egoism *may* promote actions that are very similar to, if not the same, as those actions promoted by the major ethical theories. Secondly, egoism is not as clearly counterintuitive (in terms of the acts the theory promotes) as is often supposed and at times *may* more closely fit with our intuitions about correct action than do the actions promoted by more generally accepted moral theories; Saving the women and children first intuitively appears to be the correct act, yet this is the act promoted by the egoist's doctrine (from self-interest) as opposed to the utilitarian whose doctrine advises sacrificing the women and children in favour of the more able-bodied (in order to maximise survival chances).

At this point, I will make it clear that I am not claiming that egoism always promotes acts that match our moral intuitions or that are compatible with, or intuitively superior to acts promoted by other moral theories. I am only showing that ethical egoism is worthy of consideration as a moral theory and ought not be discarded out of hand.

It might be countered that my case is quite artificial and has been carefully sculpted so as to present a desired result. This is true to a degree just as it is true of all philosophical thought experiments. However, the case presented is not unusually bizarre and is certainly possible in the real world. That aside, it is worth discussing one direct objection before moving on, namely that if it were not for the likelihood of repercussions in the lifeboat case, the egoist would appear to lack any compelling motivation to rescue any of the other passengers, unless it somehow added to her own chance of survival. Sailing off alone and leaving 19 people to die is intuitively immoral and counter to the advice of all other-regarding moral theories. Thus it might be claimed that, in so far as repercussions

reflect common beliefs about the right thing to do, it is no great surprise that this makes itself felt (in conjunction with survival) in the egoist's actions.

A case study could be created where as a matter of fiat the egoist will face no repercussions and in such a world the egoist would save at most four others. Such a world would be a strange place indeed and if we lived in such a world our notions of morality would likely be very different indeed, certainly such a world would not be based on egoist principles¹¹. In this world, however, there is no way the egoist can possibly know there is no possibility of repercussions and the act must be based on the probable outcome of the choices made. It is of course a different question as to whether an egoist looking back on the decisions would judge them right or wrong (armed with the knowledge of success/failure and what, if any, repercussions eventuated).

While, at least, it ought to be true that laws and repercussions reflect our beliefs about right and wrong it is far from clear that these laws/repercussions would be so different in a society that conformed to egoist principles. For example, the egoist might reason as follows: I want laws in place that allow the greatest probability that my interests will be maximised. In a lifeboat type scenario, given the probability that any given individual is more likely to be in the water than in the lifeboat, any rational individual would want laws that maximised the probability of that individual's survival. Before an emergency presents itself, an egoist would therefore reason that self-interest lies in having a specific set of laws (acknowledging the small possibility that such laws could go against the egoist's self-interest) that favour those in unfortunate predicaments. The egoist's self-interest during such an emergency is determined, to a degree, by the existence of those laws developed before an emergency.

It might be countered that such a claim is inconsistent or circular and attempts to hide the fact that egoism is a morally pernicious doctrine. I will try to defend egoism against these and other major criticisms and in doing so will aim to uncover the necessary axioms of a valid form of ethical egoism.

¹¹ If all were egoists then those ashore (who had people they valued on the boat) would be likely to take action.

§1.2 Egoism – Criticisms

As mentioned, egoism is most often viewed as the antithesis of morality and an antithesis that must be defeated. It therefore comes as no surprise that a vast number of objections have been formulated over the years. It would be quite beyond the scope of this project to survey all of the individual objections; however, they can be categorised into several distinct types of objection. Over the next few chapters I will examine in depth the individual criticisms that I consider to be the strongest representatives for each type of objection and in doing so will begin the development of an egoist axiom that does not fall foul of these objections.

The types of objection I will consider are:

1. Egoism is self-defeating: This objection accepts egoism as a theory of action but claims it fails to meet its own goal of attaining the best results for the egoist.
2. Definitional: This objection accepts that egoism is a theory but denies that it meets the requirements of a moral theory.
3. Contradictions: This type of objection makes the claim that egoism results in an internal or external contradiction.
4. Inconsistency/Incoherence: These objections claim that egoism places inconsistent requirements on action or that, in effect, it promotes no action.
5. Pernicious Doctrine: This objection makes the claim that egoism promotes acts that are intuitively recognisable as evil.
6. Theories of the self: This type of objection makes the claim that references to self-interest as made by egoism are flawed given that the notion of a self is problematic.

Finally, I will consider pseudo-defences. By this I mean arguments purported to be defences of egoism but which are formulated in such a manner that the theory they are supposedly defending is rendered impotent. I will briefly introduce each of these objections in the following sections.

§1.2.1 Criticism that Egoism is Self-Defeating

The method used by the egoist to solve the ‘Drowning Passengers’ dilemma was a simplified probability calculation (a type of game theory analysis). This might be thought unusual given that game theory is most often used as a tool in analysing dilemmas to the effect that egoism is shown to be self-defeating. My reason for doing this was that game theory most often presents moral dilemmas as finite non-

zero games¹². In Chapter 5 I will argue that moral dilemmas cannot be treated in this way and the conclusions drawn with regard to egoism are demonstrably incorrect.

Game theory is utilised to demonstrate that egoism is self-defeating when faced with a prisoners' dilemma. A prisoners' dilemma type situation arises when the best option for one player (or moral agent) is dependent on what another player (another moral agent) does, and the first player cannot know how the other will play¹³. A simple version of the prisoners' dilemma can be formulated as:

You and a fellow convict have been arrested on suspicion of committing a crime. The evidence against both of you is quite weak and unless the guards can elicit a confession you will both receive only relatively small jail sentences. The guards approach each of you separately and offer the following deal.

If you confess and the other prisoner does not you will be set free and the other will face eight years in jail.

If neither of you confess the evidence is sufficient that you will both face two years in jail.

If you both confess then you will both be jailed for five years.

It is usually assumed that an egoist, to maximise self-interest, must confess. The criticism against egoism is that if everyone acts according to egoism then whenever a dilemma (of the prisoners' dilemma type) is faced everyone will confess resulting in everyone being worse off. In the example above the prisoners, purportedly acting as egoists, both end up serving five years in jail. If neither had confessed they would have each served only two years. The conclusion is that according to egoism the agents would have been better off not acting as egoists. A non-self-interested co-operative theory (under game theory analysis) results in greater utility for the players than does a self-interested one. However, as will be discussed in Chapter 5, moral dilemmas cannot be accurately assessed by games

¹² I analysed the drowning passengers' dilemma as a finite non-zero game. I did so to avoid introducing unnecessary complications at this early stage and to show that even assuming game theory can be legitimately applied to finite non-zero games the outcome is not always as initially expected.

¹³ The prisoners' dilemma was first published by Albert Tucker in the early 1950's.

theory if they are treated as ordinary non-zero games; moral dilemmas cannot be expressed as single events.

When assessing a prisoners' dilemma, if the set of iterations are considered, as I will argue they must be, it is highly improbable that confession is the action promoted by self-interest. Further, as Mosely (2001, par. 4) notes, "the nature of the game pre-empts other possibilities. The sentences are fixed, the choices are fixed; whilst this applies to the two prisoners, it is not obvious that every-day life generates such limited and limiting choices." Forcing a prisoners' dilemma into the mould of a single non-zero event creates a hypothetical scenario that is too distant from reality to demonstrate anything of practical value. In conclusion I will show that, once it is accepted that ethical dilemmas do not fit the non-zero game mould, egoism is not self-defeating.

§1.2.2 Criticisms – Definitional

Definitional criticisms do not object that egoism is a self-defeating, incoherent or inconsistent theory. Instead they reject the claim that ethical egoism is a *moral* theory, for it is often assumed that an essential quality of morality is its other-regardingness. For example, Nagel (1986, p. 197) writes "[M]oral requirements have their source in the claims of other persons" and similar sentiments can be found expressed in the writing of most contemporary moral philosophers.

Egoism is often question-beggingly placed outside the moral arena; morality is defined as or presumed to be other-regarding and then egoism is dismissed as nonmoral on the grounds that it is not other-regarding. This occurs to the extent that the question 'why should I be moral?' is often conflated with 'why should I not be an egoist?' As Lemos (1971, p. 381) notes, the terms 'moral' and 'ethical' are frequently, and without argument, used to denote other-regarding actions. Circularity aside, assuming morality can somehow be defined in terms of others we are left in something of a quandary – there seems to be no overriding answer to the question 'why act from other-regarding (so called moral) rather than self-interested reasons?' As I will argue in Chapter 2, there can logically be no objective answer to this question. Further, I will argue that attempting to define morality as essentially other-regarding leads to inconsistencies and contradictions

that can only be overcome by including self-regard within the definition of morality. This being the case, egoism can be examined not as an antithesis to morality but as a competing moral theory¹⁴.

Even allowing that a theory of self-interest could be moral in nature, egoism still faces the further challenge that it fails to meet even the minimal requirement to gain membership in this definitional category. Amongst these are the requirements that, in order to be considered a moral theory, it must meet certain conditions of adequacy. Three such conditions are that a moral theory must be public, universal and, perhaps most importantly, that an agent must be motivated for the right type of reasons if it is to be claimed that the agent acted morally. Gottlieb (1996, p. 3) presents a criticism that falls within the definitional category claiming that “the motive of ethical egoism is the wrong sort of motive for a moral individual to have, and undermines any other morally good motives which such a person might have”. Gottlieb’s argument, like all other definitional criticisms, begs the question. Motives are assumed to be moral when the good of others provides the motivation and nonmoral when the motivation comes from self-interest.

Another method of classifying objects (broadly a definitional criterion) is by reference to their functions and so I think it is appropriate to consider it here¹⁵. The question ‘how ought I act?’ often arises through conflicts of interest and it is commonly accepted that one of the primary purposes of morality is conflict resolution. This leads to the criticism that egoism is not a moral theory because it is unable to perform the function of interpersonal conflict resolution. Such criticisms are usually formulated around scenarios involving the distribution of scarce resources.

Assume Φ is a scarce resource and Agent₁ and Agent₂ (both egoists) will have their interests met only by obtaining Φ .

¹⁴ It is competing in the sense that moral theories define right action, and the reason for acting, differently. It may turn out on an applied level, as moral theories become more sophisticated, that there is little difference in the actions they recommend.

¹⁵ For example, common dictionary definitions of ‘clock’ include reference to the functional criterion, namely that a clock is a time keeping apparatus. Likewise we can limit the scope of what is and is not a moral theory by inclusion of functional criteria.

Therefore the action promoted by egoist theory to each agent is that they obtain Φ for themselves, with the result that conflict ensues.

It would seem that egoism offers no means of resolving a scarce resource dilemma and the egoists are reduced to a battle of strength; in such a context, might determines right. While this criticism may well hold true where scarcity is extreme it is insufficient grounds to dismiss egoism. In Chapter 2 I will argue that if the criticism relies on extreme scarcity then the criticism is too strong, in so far as in defeating egoism other-regarding moral theories are equally defeated. However, where the scarcity is less than absolute “ethical egoism does not have to logically result in a Darwinian struggle between the strong and the weak” (Moseley, 2001). Ultimately such attacks against egoism are little more than strawman arguments.

§1.2.3.1 Criticisms – Logical (internal) Contradictions

G.E. Moore believed he had fully discredited egoism and claimed in *Principia Ethica* that a “[n]o more complete and thorough refutation of any theory could be desired.” (1962, p. 99). Moore’s refutation is formulated around the nature of ‘the Good’ itself and concludes that the egoist doctrine results in an absolute contradiction. According to Moore egoism results in the position that two things (both X and \sim X) are each the sole good, which is an absolute contradiction and the fundamental flaw in egoism. It cannot be the case that “an immense number of different things are, each of them, the sole good.” (Moore, 1962, p. 101). In short, Moore’s criticism involves the determination of what is a good. An egoist agent₁ claims that agent₁’s self-interest is, by egoism, the sole good. However, another egoist agent₂ makes exactly the same claim; agent₂’s self-interest is, by egoism, the sole good. Given that in many cases agent₁’s and agent₂’s good will be mutually exclusive we are left with the contradiction:

Agent₁’s good is the sole good, and
Agent₂’s good is the sole good, and
Agent₁’s good and agent₂’s good are mutually exclusive.

In Chapter 3 I will consider two responses to Moore’s criticism. Firstly, I will briefly consider whether showing a theory to be contradictory by way of formal logic necessarily shows that theory to be flawed, given that logic itself faces

numerous objections. Secondly, I will consider Moore's claim that 'the Good' must be nonrelational and agent neutral. Smith (2003) and Rasmussen (1999) are amongst those who have presented persuasive arguments showing that 'the Good' can be agent relative. If 'the Good' can be legitimately defined by reference to an agent then Moore's contradiction is dissolved. An "immense number of different things" are goods but each of them is only a sole good relative to a specific agent.

§1.2.3.2 Criticisms – Logical (external) Contradictions

Moore's criticism was that egoism contained an internal contradiction, effectively rendering the theory itself contradictory. However, a successful defence of egoism from the charge of internal contradiction does not save egoism from the criticism that, when applied to a particular case, it renders contradictory or inconsistent results.

If we assume that 'the Good' can be defined relative to a specific agent, recognise that good is still a Good-in-itself and, and follow Moore in thinking that recognition of a Good-in-itself obligates an agent to act, then there can be no *ad hoc* restrictions on which Good-in-itself obligates action. Thus, where Goods-in-themselves come into conflict, a contradiction still arises and the egoist is placed in a position whereby an obligation to both X and to ~X exists.

Smith (2003) attempts to defend egoism from this criticism by appealing to the psychological states of the agent, and I will argue that this defence fails on several counts. Instead, drawing on Rasmussen's (1999) account of flourishing (a version of 'the Good') from a specific standpoint I will argue that 'the Good' relative to some other agent does not obligate action for the reason that 'the Good' can only be recognised from within a specific worldview and we have no direct access to others' worldviews.

§1.2.4 Criticisms – Inconsistency / Incoherence

The criticism that egoism is inconsistent is similar in form to the criticism from contradiction but differs in one vital aspect. Whereas arguments from contradiction claim that the egoist is obligated to perform two (or more) mutually exclusive acts, arguments from inconsistency claim only that the egoist desires

that mutually exclusive acts occur. Arguments from incoherence usually follow from arguments from inconsistency. Whereas arguments from inconsistency attempt to show that the egoist desires inconsistent states of affairs to come about, arguments from incoherence attempt to show either that:

- (i) given that egoism advises inconsistent actions it in fact advises no discernable action, or
- (ii) it is not possible to act as egoism advises.

In presenting an argument similar in structure to Moore, Medlin (1957) claims that egoism results in an inescapable inconsistency in so far as any agent subscribing to egoism will find themselves in a position where they will want both A and $\sim A$ to occur and while such a position is not contradictory it is clearly inconsistent. To be inconsistent is to desire to have your cake and to desire to have eaten it, that is, it is inconsistent but logically possible to hold both desires simultaneously. If an agent equally wants A and wants $\sim A$ to occur, and wanting A and wanting $\sim A$ is a result of applying the egoist doctrine, then egoism has offered inconsistent advice on how to act, and has in effect offered no advice and is incoherent as a moral theory of actions.

Egoism is also criticised for being inconsistent in that it may recommend actions that prevent the egoist from following the egoist doctrine. Examples of this criticism are usually formulated by way of thought experiments but examples can also be found in ordinary life. In order to protect his own interests (especially in times of war) an egoist may find that it is in his best interests to join the army. In the army (in order to best protect his own life) the egoist will need to learn to work in a team and develop an attitude of following commanding officers' orders. However, in many cases following such orders (which will be for the good of the unit/mission) will be other than the actions promoted by egoism. It would seem that in such cases egoism advises that in order to promote self-interest the egoist ought not to be an egoist. The arguments from inconsistency and incoherence will be discussed in Chapter 4.

§1.2.5 Criticisms – The Parity of Moral Theories and the Nature of the Self

When discussing egoist doctrines, it is generally assumed that it makes sense to talk of the rational long-term interests of an agent. However, in doing so the egoist leaves himself open to the criticism that the theory subscribed to is a hybrid¹⁶ and that hybrid theories are (a) inconsistent and (b) inferior to pure moral theories¹⁷.

Moral theories can be either agent neutral or agent relative but not both, temporally neutral or temporally relative but again not both. A moral theory is temporally neutral if it includes the claim that Good is morally important and obligates (or gives good reason for) action, regardless of when it occurs. Conversely, a temporally relative moral theory claims that Good is relative to some specific point in time and only obligates (or gives good reason for) action at that point in time; the egoist adopts a *temporally neutral* position. A moral theory is agent neutral if it includes the claim that the Good is morally important and obligates (or gives good reason for) action for any agent, whereas an agent relative moral theory claims that good is relative to an agent and only obligates (or gives good reason for) that agent to act. The egoist adopts an *agent relative* position, in part to overcome the criticism that egoism is contradictory.

The criticism against egoism is that moral theories must have parity, they must either be completely neutral or completely relative. There can be no possible justification for the arbitrary selection of relativity (in the case of agents) and neutrality (in the case of temporal concerns) within a single moral theory. As Nagel (1970) notes, if the egoist's future self provides reasons to act (temporal neutrality), then others also provide reasons to act (agent neutrality). Likewise, Parfit (1984) assumes parity is required and claims that if the egoist has no obligation to others then future interests cannot override present interests.

Even assuming it can be shown that parity is not required by a moral theory and that we have good reason for treating agent and temporal concerns differently - and Sidgwick (1913) and Brink (1992) present convincing arguments to this effect

¹⁶ A hybrid theory is one that mixes aspects of two or more theories, for example, being absolutist in one aspect and relativistic in another.

¹⁷ I have included the criticism that hybrid theories are inferior to pure moral theories under criticisms relating to the self because it introduces the concept of presentism and temporal slices which are central to the criticisms regarding the nature of the self.

- the egoist still faces a challenge from presentism (an agent and temporally relative theory). The presentist claims that an agent ought to act so as promote the agent's immediate self-interest.

The challenge to the egoist is as follows: If the utilitarian (or any agent and temporally neutral theory of rational benevolence) must answer the egoist's question 'Why should I sacrifice my good for the good of another?' (a question to which it is assumed the utilitarian cannot present a convincing answer) then the egoist is logically compelled to answer the same question from the presentist, 'Why should I sacrifice an interest now for one in the future?' While presentism (in this pure form) is a highly implausible moral theory the challenge to those who subscribe to egoism is logically equivalent to the one presented by the egoist to those who subscribe to agent neutral theories, and it represents a challenge that must be overcome.

Assuming the presentist challenge can be overcome this is still insufficient, for a temporally neutral version of egoism faces further problems with regard to the self. As Gough (1998) points out, the egoist acting now does so to benefit not himself but some future self; the beneficiary of the act is some distant self who may have a completely different set of interests. The egoist cannot claim to be acting for the benefit of himself when he will not be that same self at the point where benefits are accrued. In effect, in acting for the benefit of some future self, the egoist is acting in an altruistic manner. Zemach (1978) argues a similar point by claiming that egoism is based on a 'howler' and that egoism is either irrational or meaningless. Zemach claims that the notion of a 'self' is confused, a matter of convention rather than fact, and consists of a series of overlapping temporal slices (depending on the criteria used) which bear only degrees of similarity to each other. The self (agent and beneficiary) to which the egoist refers cannot be clearly identified, and if the self cannot be clearly identified across time then the egoist cannot act for his long term best interests.

§1.2.6 Criticisms – Defending Egoism

Surprisingly, even those who defend egoism are far from content when they judge their defense to have been successful. After finding no arguments that successfully show egoism to be either inconsistent or incoherent, Lemos (1971, p. 392), confesses that he wished the opposite were true:

I therefore wish that there were some way to show, to the satisfaction of everyone, that egoism is incoherent or inconsistent. But I must confess that if there is I do not know what it is. I therefore hope that someone can point the way out, for I, at least, cannot remain content with the position of this paper, nor do I see how anyone who accepts non-egoism can be content with it. But then perhaps I am hoping for the impossible. And if the argument of this paper is sound that is precisely what I am hoping for.

Perhaps such sentiment is the result of the presupposition that egoism must be flawed and necessarily promotes immoral actions. Perhaps it results from the fact that most of those who take up the task of defending egoism find the theory distasteful and thus devote little space to developing its axioms. However, the result of this method is that a theory is defended that few if any would actually subscribe to.¹⁸

Even Kalin (1995) who presents perhaps the most widely known defense of egoism devotes little of his article to developing the theory. The result is that while Kalin (1995, p. 89) claims that it “makes sense to speak of egoism as a morality” given that it provides an answer to the question ‘What ought I do?’ and any theory that provides a “coherent answer to it thereby deserves to be regarded as a moral theory”, the egoism he defends ends up sounding far from attractive. Kalin’s egoism is a private morality that cannot be promulgated and lacks most of the features associated with moral theories. The egoist cannot enter into moral debate, cannot receive moral advice, cannot give moral advice, cannot establish institutions of reward and punishment, cannot teach his children morality and cannot justify his behaviour to others. As such, Kalin’s defence of egoism comes across more as a criticism and we are left wondering of what use, if any, such a private morality would be (if indeed, given Wittgenstein’s critique of private language, a private morality is even logically possible).

¹⁸ I am not sure that the existence of a theory that is valid but which no one would subscribe to presents anything more than a trivial theoretical problem.

Therefore, Kalin's defence will be treated for the most part as a criticism and it will be shown that a developed egoism need not be a private morality and an egoist can give and take moral advice, and join in moral discourse.

§1.2.7 Criticisms – Egoism is a Pernicious Doctrine

The final criticism of egoism is often considered the weakest from an analytical perspective. This opinion, however, is based on a less than complete understanding of the criticism. If egoism falters it is because it is unable to counter the challenge that it promotes acts clearly recognisable as morally wrong.

In the final chapters of Part 1 I will examine Rachels' (1974) claim that egoism is a pernicious doctrine. This claim is usually dismissed for the simple reason that it relies on a circular argument; Act Y is clearly recognisable as morally pernicious, Egoism promotes Act Y, therefore egoism is pernicious. Given that egoism is a moral theory that determines right acts it cannot be the case that it is wrong because it promotes an act that is judged as morally pernicious according to some other moral standpoint. However, such counters mistake the strength of Rachels' argument which is based on the claim that some acts are objectively recognisable as morally wrong with far more certainty than we can ever be sure about the correctness of any moral theory. In effect, Rachels is making a claim about moral facts that, if true, would clearly demonstrate that egoism is flawed. In examining Rachels' theory it will be necessary to consider the nature of moral facts. In doing so I will examine Harman's (1977) argument and in countering this position show that moral facts do indeed exist, albeit in a limited sense. This makes Rachels' criticism of egoism far more difficult to overcome.

But before such criticisms can be considered it is necessary to determine whether morality is necessarily or by definition other-regarding and, assuming the answer is no, whether egoism can meet the minimal conditions of adequacy and be defined as a moral theory.

Chapter 2 Definitional Issues

§2.0 Introduction

It is commonplace within discussions of the nature of morality to classify self-regarding actions as outside the concept of morality. Gert (2002) denies that morality governs self-regarding behaviour (ch. 2, para. 18). Gert is not alone in such views; according to Rogers (1977) most contemporary philosophers would insist that morality is essentially other-regarding. Generally, philosophers who adhere to this opinion subscribe to what is referred to as the self-other model; the view that “an action has no moral worth unless it benefits others – and not even then, unless it is motivated by altruism rather than selfishness” (Rogers, 1977, p. 1). The very first objection to egoism is not that egoism is a flawed moral theory but that it is not even a moral theory.

As Rogers (1977, p. 2) notes, it is not so much that theories of self-interest (notably egoism) are “demonstrably false ... but rather that [egoism] contradicts one of our most deeply held dogmas about morality”, that the very concept of morality involves the promotion of, or regard for, the good of others. If such a concept is accepted as part of the definition of morality then egoism, or indeed any theory which evaluates right action in terms of maximisation of self-interest, is predefined as something other than a moral theory¹⁹. As Hart (1978, pp. 3-6) notes about the concept of punishment, to apply a definitional stop and claim that egoism is not a moral theory simply because it falls outside of some commonly accepted definition is the wrong type of answer and unlikely to convince an advocate of egoism that they are not following a moral doctrine. The very investigation at hand is whether egoism is a moral doctrine and this question cannot be answered by reference to a definition alone.

Rogers goes on to cite several contemporary authors from across the field of ethics who explicitly accept the self-other model and build the concept of other-regardingness into the definition of morality. I will begin this chapter by briefly examining the positions extolled by three ethical theorists each of whom,

¹⁹ As is indicated it is not denied that egoism provides prudential reasons for action only that these reasons are not moral reasons.

according to Rogers, embraces the self-other model. I will then consider what justification and reasoning might warrant the exclusion of self-regarding interests from the moral domain.

§2.1 Peter Singer's Utilitarianism

According to Singer, “[t]he ethical life is the most fundamental alternative to the conventional pursuit of self-interest” (Singer, 1997, p. vii). It should be made clear that Singer (1997, p. 23-24) is not claiming that self-interest and morality are logically exclusive, but rather (1977, p. 4-9) that the gratification of immediate self-interest, for the most part greed and selfishness, is the antithesis of morality. Singer allows that an alternative enlightened form of self-interest where the ‘ethically reflective life is also the good life for the person leading it’ may provide a solution to the self-other conflict. Even so, the reason for acting is still the good of others, and the ‘good life (for me)’ is defined in terms of living an ‘(other-regarding) good life’.

Further, utilitarianism itself is not entirely other-regarding. The (self-interested) preferences of an agent are given equal weight with the like interests of every other person affected by a moral decision. Self-preferences are reduced to one amongst the many rather than being totally excluded, as stated by Mill (1963, p. 218) “each to count for one [including the self] and none to count for more than one”. Assuming the possibility of a moral decision where no other person is affected by the agent’s decision (and for Singer this means no other sentient creature), then right action (maximization of interests) and self-interest will necessarily coincide. However, even in this rare case, the reason for acting is not out of my own self-interest but to maximise interests, giving equal weight to all equal interests of all those affected by the decision, which in this case just happens to be limited to me alone.

In the majority of moral cases many others will be involved and my own self-interest, while given equal weighting, plays only a very small part. In any moral decision, where numerous people have an interest in the outcome, the act promoted by utilitarianism (that is the act that maximises interests) is likely to be the same regardless of whether or not my specific self-interests are counted. While

utilitarianism does make a small place for self-interest it denies it any primacy. The reason for acting is the maximization of, primarily, others' interests even when this may mean sacrificing self-interest.

§2.2 Rights and Virtues

The extent to which other-regardingness has primacy within virtue and deontological theories is even more pronounced. Aristotle allowed for the inclusion of self-regarding virtues, such as moderation, within his moral theory. However, Thomas Nagel, a Neo-Kantian, denies Aristotle's position that "The moral life is defined in terms of the good life" (1986, p. 195), which might include self-regarding virtues, and instead claims that "moral requirements have their source in claims of other persons" (1986, p. 197). According to Nagel if an action is not for the good of another it is not a moral requirement. Likewise, Lawrence Blum, a virtue theorist, argues that for a trait to have any moral value it must be directed toward the good of another:

Basically what makes the altruistic emotion morally good is that its object is the weal of another person. Why it is of moral value to have sympathy, compassion, or concern for someone is that one is thereby concerned for the good – the weal and woe – of another person (Blum, 1966, p. 163).

Some virtue theorists admit that virtue ethics is at least formally egoistic and that developing the virtues is of benefit to the possessor - not that possessing the virtues guarantees flourishing but that it is the best bet for having a flourishing life (Hursthouse, 1999, p. 173). However, the virtuous person is expected to promote flourishing and to act in accordance with the virtues even where to do so is quite clearly detrimental to the agent's own flourishing. While the ultimate reason for possessing the virtues may be some form of enlightened self-interest the reason for acting virtuously in some specific context is the good of others.

§2.3 Appeals to the Best Explanation

If we resolve to define morality as essentially other-regarding or necessarily requiring other-regarding sacrifices then we are presented with something of a quandary; namely 'why be moral?'. If morality means, by definition, 'acting for the good of others' then why should I preference reasons for acting so as to

benefit others (i.e. moral reasons) over reasons for acting to benefit the self (i.e. prudential reasons)?

As I see it, there are only two methods by which to argue for the truth of the proposition that reasons from an other-regarding standpoint offer superior justification for acting than do reasons from self-interest, religious conviction, financial obligation or aesthetic notion.²⁰ Firstly, it might be argued that self-regarding reasons are fundamentally flawed as reasons for acting. Prudential reasons for actions are, in effect, only pseudo-reasons because they fail to meet the minimal requirements of being valid, noncontradictory, and consistent.²¹ As long as we accept that valid reasons must meet minimum criteria of adequacy (whatever those criteria might be) and that reasons from self-interest fail to meet these criteria then, by default, other-regarding reasons (so long as they meet the criteria of adequacy) are superior. Egoistic reasoning is criticised on the grounds that such reasons fail to meet one, or more, of the basic criteria of adequacy. I will briefly examine these criticisms later in this chapter, and will argue that they are not successful in showing that other-regarding reasons are superior by default.

Secondly, it might be argued that, while self-interest does provide reasons to act, other-regarding reasons provide superior and overriding reasons for action. However, such arguments are conceptually confused. For any given proposition there will be numerous competing hypotheses H_1 , H_2 , H_3 , that equally explain the proposition and all meet the minimum requirements of adequacy. For example, given that we want to explain the existence of the universe we could invoke a scientific hypothesis (H_1) or a religious, creationist explanation (H_2)²². How then are we supposed to choose between H_1 and H_2 and label one of the two hypotheses superior? If we are to avoid circularity then we cannot choose between the two hypotheses by reference to either of the existing hypotheses. While it is necessarily true that, according to scientific criteria for a best explanation, H_1 (the

²⁰ For the sake of simplicity I will limit the discussion to consideration of other-regarding reasons and self-regarding reasons for action.

²¹ This is in no way meant to be a list of the minimum requirements of a good reason. The criteria given are merely being used to indicate the type of attack that might be used to discredit prudential reasoning.

²² I am oversimplifying in assuming only single scientific and religious explanations exist. However, this has no impact on the arguments that follow.

scientific explanation) is the better explanation it is equally true that, according to the religious criteria for a best explanation, H_2 (the religious explanation) is superior²³. It would seem that any noncircular selection of H_x as the best or, at least, most probable explanation requires an independent criterion by which to rank the competing hypotheses. Without such a criterion, as Sayre-McCord (1988, p. 278) notes, we will then not be saying that one explanation really is better than another, only that we have societal or personal preferences regarding explanations. Those inclined toward scientific beliefs will preference scientific explanation; those inclined towards religious beliefs will preference religion. Likewise, without some form of overriding and independent criterion, labelling other-regarding theories of action as superior to self-regarding reasons for action is nothing more than societal or personal preference and holds no compelling force.

The alternative to dismissing the claim that some valid explanation is superior to another valid explanation is to hold to the proposition that there is an overriding criterion (with the property of being superior) by which the best explanation can be determined. This might be accomplished if we could locate some unchallengeable criterion to apply but if no such criterion exists, as seems likely to be the case given that such criteria are constructs, then any claim that X represents the criteria of a best explanation is logically incoherent.

Let E_1 be any set of criteria by which to determine a best explanation.

Let E_2 be any second set of (different) criteria by which to determine a best explanation.

Sets E_1 and E_2 are mutually exclusive.

Just as it is invalid to select H_1 or H_2 as a superior explanation by reference to H_1 or H_2 it is equally invalid, for the same reasons, to select E_1 or E_2 as a superior set of criteria by reference to either of the existing criteria (E_1 or E_2). Such an answer would be circular and show only that the set of evaluative criteria E_x was

²³ Scientists often make the claim that scientific explanation is superior because it meets the scientific community criteria of being testable by experiment and able to produce predictions about the world. The religious community claims the religious explanation is superior because it explains supernatural phenomena (that by its very nature is not testable by experiment.)

preferable because it met the criteria of E_x . The selective criteria are then only preferred, not superior. The only way to demonstrate that E_1 or E_2 is a superior set of evaluative criteria is by reference to a third (and ultimately superior) set of evaluative criteria.

Let F_1 be a higher level set of evaluative criteria that has the property of being superior.

If F_1 provides the means to judge between the superiority of not only E_1 and E_2 , but also between H_1 and H_2 then E_1 and E_2 are rendered obsolete. If we have an ultimate form of explanatory criterion then we have no need to appeal to E_1 or E_2 . However, the claim that F_1 is superior to E_1 and E_2 is itself in need of justification and any attempt at explanation leads to a *reductio ad Absurdum*.

1. Showing F_1 to be a superior set of criteria cannot be accomplished by reference to F_1 as such an explanation would be circular.
2. F_1 cannot be shown as superior by reference to E_1 or E_2 . These are already statements of preferable selective criteria that differ from F_1 . Reference to E_1/E_2 would necessarily result in E_1/E_2 being selected as the preferable set of explanatory criteria.

I am not dismissing the importance of explanatory criteria (and I shall return to a discussion of explanations in Chapters 5 and 6). My claim is only that given two opposing explanations (science versus religion, or self-regarding versus other-regarding reasons for action) it is logically impossible to show that one of the explanations is superior to the other where both offer valid explanations from within their paradigmatic standpoints. Those who hold tightly to the religious criteria will never be convinced by scientific explanation where the explanation differs from their own, and vice versa.

It might be claimed that there is some significant difference between the type of explanations required for things in the world (and the conflict between the various available explanations) and the type of explanation required for theories of how we ought to act. This, however, does not appear to be the case. The conflict between self and other-regarding acts is demonstrated in the Ring of Gyges:

Gyges, a shepherd, discovers a ring that has the power to make him invisible. Gyges travels to the capital, seduces the queen, and with the power to move undetected conspires with the queen against her husband, slays the king and takes control of the city (Lydia). (Plato, *Republic*, 359d-360b)

Glaucon claims that whoever possessed such a ring, and had the ability to move undetected and with no possibility of retribution, would soon submit to self-interest and hedonistic impulses.²⁴ The dilemma faced by Gyges is ‘why, free from any possibility of retribution, ought we act from other-regarding motivations rather than self-interest?’ The basic conflict between other-regarding reasons to act and self-interested reasons to act can be phrased, for any individual act:

- (i) An agent has reason to act for the good of others. Gyges *could* have acted, and had reason to act, in an other-regarding manner.
- (ii) An agent has reason to act for the agent’s own good and, in doing so, acts other than for the good of others. Gyges had reason to act as he did, out of self-interest.

Rationale can be provided for both types of reasons and, as has been claimed, where the reasons are mutually exclusive there can be no logically compelling reason as to why preference should be given to other-regarding reasons over self-interested reasons. It is, however, often claimed that reasons from an other-regarding standpoint are normatively more important than reasons of any other form. Copp (1997) presents the following to show the inherent flaw in claiming that other-regarding reasons are overriding:

1. If the claim that ‘S’ (other-regarding morality) is normatively the most important standpoint then it must be judged from an authoritative normative standpoint ‘R’ that yields the verdict that ‘S’ is normatively the most important standpoint.
2. If ‘R’ is to be authoritative then either ‘R’ is of equal importance or is normatively more important than ‘S’.
3. ‘R’ cannot be of equal importance to ‘S’, (as was discussed earlier in this section) for no set of criteria or theory can be judged supreme from within

²⁴ Such dilemmas have been examined through the mediums of stories and films, for instance; H.G.Wells’ *Invisible Man*, Rubin/Ramis’s *Groundhog Day*, and Tolkien’s *Lord of The Rings* in which Gandalf (among others) turns down the One Ring of Power recognising that although he would attempt to use it to do great good he would through it, out of his own self-interest, do great evil. (Through Gandalf’s statement Tolkien, like the philosophers discussed earlier in the chapter, implies that the nature of morality is other-regarding).

its own standpoint. 'S' will always be supreme when judged from within 'S' just as 'I' (reasons from self-interest) will always be supreme when judged from 'I'.

4. Therefore (by 2 & 3) 'R' must be normatively more important than 'S'.
5. Therefore (by 1 & 4) 'S' cannot be normatively the most important standpoint.

Copp (1997 p. 103) sums up the argument: "The incoherence can be displayed in two sentences: The claim that a standard 'S' had the property of supremacy is the claim that it is normatively the most important standard as assessed in terms of some other standard 'R' which is normatively the most important standard. But only one standard could be normatively the most important".

Why be moral (act for other-regarding reasons), rather than act out of self-interest? There can be no objective answer to this question. At best it might be answered that there is a societal preference for acting from other-regarding reasons, whereas the question, "why act from self- rather than other-regarding reasons", has a compelling, if circular, answer; because it is in my interest to do so. If it cannot be shown that other-regarding reasons are normatively superior to reasons from self-interest then dismissing theories of action based on self-interest requires demonstrating that such theories are in some way flawed. Whether this is possible will be considered throughout Part 1 of this thesis.

§2.4 Defining Morality; Self versus Other

Assuming that reasons of self-interest are valid and given that there is no overriding reason to act from other-regarding rather than self-regarding reasons I am interested in whether it makes any sense to limit the definition of morality to essentially other-regarding action. Defining theories of self-interest as outside morality, and thus viewing reasons based on those theories as nonmoral results in many seemingly moral questions having no moral solution; 'How ought I to act in this context given that no one else is affected?' is not a moral question on this view despite an intuitive feeling to the contrary.

§2.4.1 The Lack of Self-Guidance

If moral theory offers no guidance where others are not involved then absurd consequences follow, consequences that despite intuitions to the contrary we are unable to label as immoral. Consider the following scenario;

On a spaceflight you are knocked off course and crash land on a planet uninhabited by any form of sentient life. Further, given the remote part of the galaxy this planet is located in it will never be of any use to any other civilisation. The planet has an abundance of vegetation and could support you indefinitely. However, you see no purpose in living a solitary lifestyle but are not the type of person to simply and quickly end your own life. Instead, you decide to destroy every living thing on the planet (and you have in your spaceship the destructive capacity to allow you to complete this task). Just before commencing this rampage of destruction you are stuck by the thought ‘is it morally permissible for me to follow this course of action?’

This seems like a moral question yet, according to Gert’s and many others’ (other-regarding) definition of morality there is nothing *morally* wrong with this course of action. The act of destroying all life on the planet is of no moral significance for the simple reason it will affect no others²⁵.

Rogers (1997, p. 2) considers a similar Crusoe example. A person is shipwrecked on a small island with the means to survive, but alone and with no hope of rescue. Rogers questions whether we really want to say there is no moral difference between (a) the strandeer who strives to improve their life, despite the difficult circumstances; building shelter, using sustainable food sources, cultivating where possible, and spending their time as constructively as circumstances will allow, (b) the strandeer (such as in my example) who wages a private and destructive war against the environment or, (c) the strandeer who is lazy and simply gathers food when necessary and does the bare minimum required to survive.

It would appear that the strandeer who strives towards self-improvement and an improvement in living conditions is exhibiting virtue yet, as Rogers (1997, p. 3) points out, on the self-other model such “virtue has no specifically moral merit”.

²⁵ I am assuming ‘others’ is limited to, at the most, sentient beings. However, it may be that our obligations do not just concern other agents, but include other living things too (and perhaps even all other things). The issue is not *just* about the extent of the moral community in the sense of a community of agents. Even so, the example does show that there are moral issues that aren’t other-regarding in the ordinary sense.

If morality is defined in terms of other-regardingness then, where no others are involved, we must deny any moral distinction between someone who acts to better themselves and someone who is self-destructive. It is not that the self-destructive person acts immorally on this view; I am not defining morality as altruistic and then claiming that any act which is not other-regarding is immoral. The claim is simply that if morality is defined as essentially other-regarding then the self-destructive acts in the example have no moral status whatsoever; all of the acts (a,b,c) have no moral significance.

While nothing discussed in this section shows any logical problem in defining morality as other-regarding we can question whether a morality that fails to answer the apparent moral question ‘how ought I to act? (in a context where no others are involved)’ is really the best definition available.

The need for self-guidance and reasoned answers to how I ought to act when others are not affected by my actions provides strong argument for the inclusion of acts from self-interest within the definition of morality. If there is good reason to exclude self-interest from the definition of morality then there must be some substantial difference between other- and self-regarding reasons. It is to those differences that I will now turn my attention.

§2.4.2 A Question of (Moral) Effort

It is sometimes argued that the difference between other-regarding and self-regarding actions is that other-regarding actions require special effort, (sometimes question-beggingly referred to as moral effort) whereas it is considered easy to act out of self-interest. Such claims are demonstrably false. Rogers (1997, p. 3) notes “parents seem inclined to put their children’s good before their own” and there are innumerable examples of parents putting their lives at risk to enter hazardous environments to rescue trapped children. As Singer (1997, p. 107) notes “the readiness of parents to put the interests of their children ahead of their own interests is a striking counter-example” to the claim that self-interest is always the easiest option. In general everyday circumstances, and not just in cases of

emergency, it seems unlikely that most parents would find it easy to act out of self-interest and spend money on holidays at the expense of their children²⁶.

Unless 'self' is somehow expanded to encompass all those that are close to us then, in some limited contexts (perhaps only those involving their offspring, or the larger group an agent values), the inclination and therefore the easier choice seems towards other-regarding and away from self-regarding actions. In many cases it requires far greater effort to act from self-interest. For example, parents are naturally inclined to care for others, their own offspring, despite any risks to themselves. Parents acting for the sake of their children, at the expense of self-interest, casts doubt on the claim that sacrificing self-interest requires special effort. Acting out of self-interest is not, by default, the easy option.

Further, unless self-regarding actions are defined in terms of simple, in the moment, hedonistic pleasures then people often have great difficulty acting out of rational self-interest. It is far easier to give in to a craving (for example smoking) than to recognise a self-interest (long term health), and resist. If other-regarding actions hold some special status it is not because they, as a rule, require a greater effort on the part of an agent. Perhaps then it is not so much that the motivation for other-regarding actions is necessarily harder but that it differs in kind from self-interested motivation.

That an ethical theory defines right by a set of criteria does not necessarily imply that we will be motivated by those criteria. However, it does appear that, if an agent is searching for an answer to the question 'how ought I act?', the answer to that question will provide some form of motivation.

²⁶ I am not claiming that parents always put their children's interests first. Instead, my point is that generally children are fed (more than the absolute minimum), clothed (again more than the minimum), provided with school materials etc when parents could have acted otherwise.

§2.4.3 The Myth of Pure Motivation

If we allow, in principle, that both self- and other-regarding interests could provide moral motivation, then the following objection may be raised:

- (i) Motivation toward self- and other-regarding actions may lead to an agent being motivated to act in mutually exclusive ways.
- (ii) Other-regarding motivations will be impure. That is, they will be tainted by self-interested motivations.

The first of these objections is not really related to the inclusion or exclusion of self-interest from morality. If there are situations in which self and other-regarding interests will motivate the agent in mutually exclusive ways then potentially we do have a problem. However, this in itself does not give reasons for excluding self-interest from the definition of morality. The problem could just as easily be overcome by the exclusion of other-regarding interests. This could even be considered a preferable answer to the criticism given that other-regarding concern for two or more agents might similarly have a mutually exclusive nature. Whether there is a deeper problem involving contradiction and inconsistency will be discussed in Chapters three and four.

The second objection is simply that moral motivation requires purity; there cannot be multiple motivations. If self and other-regarding motivations are both moral motivations then it may turn out that agents have ‘one thought too many’. The purity of motivation, for the good of others, is tainted when the agent further considers whether the act is in the agent’s self-interest. The notion of enlightened self-interest often surfaces at this point, namely acting in an other-regarding manner will ultimately be in the agent’s self-interest. As Rogers notes, “Philosophers who stress motivational purity sometimes observe that when we deal justly and benevolently with others, we tend to do best by ourselves” (1997, p. 12). The point is that an agent’s enlightened self-interest is furthered by not directly considering that self-interest (that is, by not allowing that self-interest to be a motivational factor). If we act generously and honestly toward others, motivated by a desire for their good, then it is probable we will be well accepted in the community and be able to form meaningful relationships. Ultimately this will *probably* be in our own self-interest in as much as the benevolent course of

action, all things considered and in most contexts, is likely to have the greatest chance of producing consequences that are in the agent's enlightened self-interest.

But how is this supposed to work? Presumably, if an agent follows the virtuous life he does so with the knowledge that it is the best way to act and the way most likely to result in a good life for the agent. However, purity of motivation requires that the agent be in no way motivated by that knowledge. The motivation must be solely for the good of others. This recommendation, somewhat bizarrely, tells us "that morality will further our interest, but in order to act for our interest, we cannot act for our interest" (Rogers, 1997, p. 12). I doubt that pure motivation is even logically possible. As Rogers remarks, "Requiring unself-interested concern for other people borders on the oxymoronic: It tells one to act toward them in a concerned – i.e., interested – way, but not to be motivated by that concern, i.e. interest" (1997, p. 13). It seems absurd that all thoughts that a specific action is ultimately in my own self-interest must be struck from my mind else the exact same action loses all of its moral force.

§2.4.4 Inconsistencies in Defining Morality as Essentially Other-Regarding.

Defining morality as other-regarding leads to numerous inescapable inconsistencies. The three inconsistencies that I will examine are the Darwinian double standard, logical inconsistency and, discrimination against the self.

The Darwinian double standard occurs when claims are made that we have an inclination towards other-regarding morality and at the same time denying any worth to inclinations from self-interest because we are inclined toward them. Singer (1997, p. 124) concludes his chapter on selfishness with the claim:

Human beings often are selfish, but our biology does not force us to be so. It leads us, on the contrary, to care for our offspring, our wider kin, and, in certain circumstances, for larger groups too.

Are the other-focused actions described by Singer to be classified as moral acts? They meet the criteria of moral acts in that the acts are other-regarding and they also meet other criteria commonly associated with definitions of morality. For example, they are likely to reduce harm or promote flourishing, and could easily

be part of a public system. If Singer is correct and our biological makeup leads us to care for others then it seems that we can extract the implicit generalisation:

- (i) Others form a moral end, and
- (ii) We (*the self*) are inclined toward this end.

Yet the self-other model embraces the inconsistency that:

- (i) The self cannot form a moral end, and
- (ii) We (*the self*) are inclined toward this end.

According to Rogers this inconsistency involves the claim that, (a) as society evolves virtue will win out and we will be inclined toward morality, *and* (b) denying that self-interest has any moral worth because we are inclined toward it.

Secondly, there is a logical inconsistency in the self-other model. Dewey (1908, p. 364) points out that consistency requires that the self either is or is not a moral end. However, the self-other model leads to a contradiction where the exact same act that results in the exact same moral end has a different moral status dependent solely on the agent carrying out the act.

Consider the following scenario:

I (Agent_{me}) through no fault of my own have insufficient food.

Presumably the act (X) of providing me (Agent_{me}) with food, where the act does not impinge on any other agent, is a good one. Indeed where some agent (Agent_{other}) carries out act (X) to benefit Agent_{me} the act is defined by most philosophers as a moral one.

Yet, when the agent and beneficiary are the same (that is where Agent_{me} acts to provide Agent_{me} with food) the act (X) is denied any moral status.

Assuming no other conflicting moral obligations; if providing food for those in need is good then providing food for the hungry (whoever the hungry might be) is a good act and whoever provides the food acts rightly. If the beneficiary and the provider happen to coincide then an act of self-interest must have moral standing. Adherence to the self-other model leads to an inconsistency. Removing this inconsistency requires, according to Rogers (1997, p. 4), either:

(i) Reconceiving our own self-interest in a manner that does not question-beggingly build immoralism right in, or

(ii) Banishing *all* self-interest from the moral domain, *others included*.

The final inconsistency I will look at, that of banishing self-interest from the domain of moral theory, can be explicated by the statement ‘What’s so different about me?’. One of the criticisms often leveled against egoism is that it fails to show that the self is different in any morally relevant way, such that a line of demarcation can be drawn between the self and others in assessing the legitimate beneficiaries of moral actions. It is commonly accepted that to exclude any group from the moral community, unless they can be shown to differ in a morally significant manner, is inconsistent.

There are factual (sexual) differences between men and women, and factual (racial) differences between Europeans and Africans. However, if these groups are to be treated differently then it needs to be shown how these factual differences make a moral difference – and it seems that no such moral difference exists. As Singer (1976, p. 150) points out “from the mere fact that a person is black, or a woman, we cannot infer anything else about that person”. Thus egoism is charged with being flawed as it cannot show any moral difference between the self and others and thus has no reason to privilege the self over others. Yet, those who promote the other-regarding model of morality are guilty of this exact form of discrimination; they privilege others over the self.

§2.4.5 Admiration of Self-interest

I think enough has been said to cast considerable doubt on the practicality of defining morality as other-regarding (or for excluding theories of self-interest); it is question-begging, logically flawed and inconsistent. Further, it seems there are cases where we intuitively believe that actions that do not affect others have moral implications (such as the Crusoe example §2.4.1). Rogers likewise argues that we do intuitively assign moral worth to actions taken out of self-interest and presents the following example:

Consider the case of a certain Vietnam veteran whose contact with a grenade during the war has left him permanently confined to a wheelchair.

Upon returning to the U.S., he decides to enrol in architectural school to pursue a lifelong ambition. Each school day for four years he wheels himself out to the bus stop at 7 a.m., takes an hour long bus ride and then wheels himself, rain or shine, around a large, hilly campus going from classes to labs to the library. At 7 p.m. he reboards the bus, goes home, fixes himself dinner, studies as long as he can, and goes to sleep. He eventually graduates top of his class and becomes a highly successful architect. (1997, p. 6)

Rogers claims that the veteran in the story would be widely praised, due to his moral virtue, despite the fact that his actions were out of self-interest. While I generally agree with the point made, Rogers doesn't spend much time examining the case and perhaps reaches her claims a little prematurely. Would we admire the same veteran, completing the same course in the same manner (and presumably showing an equal amount of courage, determination and integrity) but who was a thoroughly unlikeable, cynical and sarcastic person? Perhaps one of the reasons we would admire the veteran is because, in acting as he does and in the way he does, the veteran provides a role model. If we admire the veteran because he acts as a good role model then the moral attributes we are admiring take on an 'other-focused' aspect; we admire the veteran because he provides a role model *for others*. However, perhaps one of the reasons we admire the veteran is because he acts for himself without impacting on our own self-interest. The veteran's bravery and courage benefit himself and he is a good role model for others because he demonstrates how to act out of self-interest, that is how to be a good egoist. Why the veteran might be admired is open to debate. However, I agree with Rogers that cases such as this demonstrate that "to assume in advance, that a person's efforts at self-realization, or flourishing, must be at variance with morality, is manifestly false (Rogers, 1997, p. 6).

Rogers offers a second example that focuses on our intuition that certain acts have moral implications, and more clearly demonstrates that at times we *do* want to include actions directed toward the self within the context of morality – namely when we consider those acts to have moral implications.

A student cheats on a final exam. The exam grades are not scaled and so the student's results can have no effect on others in the class. The exam plays no part in the career the student will eventually follow and the student does not cheat on any exams where the knowledge is vital to his

chosen career. The student's lecturer is aware of his cheating but is unconcerned.

Further none of the other students know of the cheating, nor will the student's act become known to anyone other than the lecturer, so the cheating student will not become an anti-role model. (1997, p. 6)

Retaining the definition of morality as other-regarding results in being unable to say that the student's action had any moral implication; the student's actions did not affect any one else²⁷. Claiming that the student's action had moral implications results in being unable to exclude self-interest from moral consideration. The principle involved in the 'cheating' scenario is that honesty is sacrificed. To be dishonest is to distort the truth, and to misrepresent facts. The important point is that the facts are distorted, not to whom they are distorted²⁸. As Rogers correctly notes (1997, p. 8) "The question of the beneficiary of one's virtue is a separate matter from whether one has and exercises that virtue". It may be countered that deception of another is a greater wrong and does a greater harm than self-deception. This may be the case but I am not concerned here with the degree of harm. All things considered, killing is morally worse than stealing but they are both moral concerns. Likewise, all things considered, harming others may be worse than harming the self but this is not a valid reason for excluding the self from moral consideration.

It may well be that other supposed virtues, for example generosity, are best explained by reference to others. I have so far limited my discussion only to honesty, as honesty is necessary for the formulation of best explanations. However, even if it could be shown that some specific virtue is solely other-regarding this would be insufficient for barring the self from the definition of morality as a whole.

So far my claim is that, although others may provide reasons to act, the self may equally provide reasons to act, and while the good of others may provide us with

²⁷ I am assuming the students are not ranked, or that the ranking is of no relevance. It would be a different story if by cheating a student gained a higher ranking that somehow adversely affected another student.

²⁸ A discussion of honesty and its importance in formulating best explanation will be discussed in the concluding sections of Chapter 5 along with why (in most cases) honesty is in an agent's self-interest.

moral motivation, consistency requires accepting the claim that our own good equally provides moral motivation. As has been shown, separating the self and others within the domain of morality leads to logical inconsistencies. However, while self-interest in itself cannot legitimately be excluded from the definition of morality there may be good reason for excluding theories of self-interest from the set of moral theories; namely that theories of self-interest defeat the function of morality. If the function of any item Y is that it performs the function X then, in most cases, something that cannot perform the function X is not a Y²⁹. If it can be shown that an essential function of morality is X then any theory Y that is incapable of performing that function is clearly not a moral theory.

§2.5 Conflict Resolution

It is generally accepted, that “[t]he function of morality or ethics is to rationally adjudicate disputes so that societal relationships can endure” Duska (1999, p. 27). Theories of self-interest are criticised on the grounds that they offer no system for the resolution of conflict, and since they fail to perform this essential function of morality are not moral theories. Working on the reasonable assumption that those evoking a moral code are looking for ways to solve moral dilemmas (and that such dilemmas often involve conflicts of interest) then the criticism that theories of self-interest fail in this regard (if the criticism is valid) would appear to be good grounds for excluding theories of self-interest from the class of moral theories.

The argument for the exclusion of theories of self-interest might be formulated as follows:

- P1) Conflict resolution is an essential function of morality.
- P2) Theories of self-interest are unable to resolve conflicts of interest.
- C) Theories of self-interest are not moral theories.

P2 appears to be the weakest premise for clearly it is not the case that egoism is unable to solve any conflicts of interest. For example, if some resource (R) exists and both Agent₁ and Agent₂ have an interest in R then while a conflict of interest appears to be present it is not the case that a Darwinian struggle will necessarily

²⁹ I say ‘in most cases’ because a broken Y is still a Y (of some sort). Accepting that the function of a clock is to keep time does not mean that a broken clock is suddenly rendered a non-clock. However, we can say that something that was never designed to, and never has the ability to, keep time is not a clock.

follow. For either agent to monopolise R is not in that agent's best interest given that both agents also have an interest in avoiding unnecessary conflict. In such cases co-operation, not conflict is promoted by egoism. Every member of any society has an interest in gaining the necessities for survival³⁰. However, rational self-interest lies in gaining a share of that resource, not in monopolizing the entire resource, for to do otherwise would be to unnecessarily create enemies, which can hardly be considered in an agent's best interest. P2 is clearly false if it means egoism is unable to resolve *any* conflicts of interest, given that the egoist has an interest in avoiding conflicts.

Furthermore, there doesn't appear to be a problem where there is a shortage of resources such that it is not possible for everyone to gain a sufficient share. It is in each agent's interest to ensure they have as good as possible a chance of receiving their share of an insufficient resource, and they ought to co-operate with others in any scheme that ensures they are not discriminated against in the division of resources. The egoist also has an interest in ensuring that when they become available any resources the egoist gains are not at risk from those who miss out. It is therefore in each individual agent's interest to ensure there are procedures in place to make sure that those who miss out are controlled. Although the agent is fully aware that they could be the one who misses out, and acknowledges that if this occurs then all interest in having such controlling procedures in place will vanish, at the point prior to distribution of resources the egoist's interest lies in ensuring such procedures are established³¹.

Where there does appear to be a problem is in the case of extreme scarcity. Where R is a scarce resource such that only Agent₁ or Agent₂ (but not both) may acquire R and both Agent₁ and Agent₂ have a rational interest in acquiring R, egoism advises both Agent₁ and Agent₂ that they ought to obtain R, and conflict ensues. Consider the following case:

³⁰ No doubt most have an interest in gaining more than the minimum. However, to ensure the minimum for survival is still a primary self-interest.

³¹ While it is possible to create scenarios where an individual need not co-operate (for whatever reason) I am interested, at this point, only in the average agent and demonstrating that, in most cases, theories of self-interest can resolve conflicts of interest.

An accident occurs in an experimental laboratory and four technicians are infected with a deadly virus. The four technicians are isolated before the virus can spread. The good news is that there is an antidote to the virus; the bad news is that only a single dose is available (and it will not be possible to obtain additional antidotes prior to the virus killing the infected technicians). The lab doctors are unable to decide who to vaccinate and so they place the antidote in the isolation room and suggest the technicians decide among themselves who gets the cure.

The antidote is an extremely scarce resource (R) and all of the technicians (Agent₁₋₄) have a rational interest in obtaining the antidote³². It would seem that the recommendation of egoism is that each agent ought to obtain the antidote for themselves as sharing is not an option. The result is a conflict between the agents, and the outcome probably determined by the survival of the fittest.

However, unless some guarantee of success exists, the egoist will not want to enter into a fight that may result not just in death (as they fail to obtain the antidote) but in spending their last few hours of life in a great deal of pain. Thus, it might be countered that the best interest of the egoist does not lie in getting the antidote, but in ensuring they have the best possible chance of obtaining the antidote. For example, the egoist could agree to a decision being reached through the drawing of straws. However, while it might well be in the egoist's interest to participate in playing the game, as soon as the game is over the best interest of the losers remain in getting the antidote and they (the majority) have no interest in honoring the outcome of the game. Peace only lasts while the game is in progress, and then conflict returns.

Egoism seems to run into a serious problem when attempting to resolve conflicts that arise from extreme scarcity, and are such that the majority will miss out. The initial argument, reformulated in terms of conflicts of interest regarding extreme scarcity, is as follows:

- P1) Resolution of conflicts of interest (in cases of extreme scarcity) is an essential function of morality.
- P2) Egoism is unable to resolve conflicts of interest in cases of extreme scarcity.
- C) Egoism is not a moral theory.

³² There may be some debate around the nature of rational self-interest. However, if anything is an example of rational self-interest it would seem that an individual agent's survival would fit the bill.

However, in this formulation the requirement of P1 is too demanding. For while moral theories in general do not promote conflict, neither do they offer a solution to conflicts involving extreme scarcity. It is perhaps even a strength of egoism that it recognises some conflicts as irresolvable for to deny this would present an overly simple view of moral life.

Reconsider the infected technician's case from a utilitarian perspective. Unsurprisingly, that one survives and three die (a utility of -2) is better than if four die (a utility of -4) and so, by utilitarianism it is better if someone gets the antidote. Yet, unless utilitarianism offers some method of determining who gets the antidote then no practical solution has been offered. An appeal might be made to 'who offers the most to society' but there is unlikely to be much difference between four technicians working similar jobs; one might have a large family, one might do a lot of charity work, one might be working on some major project that will benefit..., it seems an impossible task to objectively rank such values and each individual is likely to think their effort ought to be ranked as more beneficial. Assuming there are no willing martyrs amongst the group, all utilitarianism offers is what everyone (including the egoist) is already aware of; (i) It is better if someone lives than if everyone dies, and (ii) It would be best to avoid conflict if possible (conflict maximises neither the individual nor the majority interest). The utilitarian, like the egoist, might agree to a drawing of straws. However, once the game is over it is difficult to see how honoring the result maximises interests (given that the majority, in honoring the agreement, are going to die). If the question 'what ought we do now?' is asked (after the game of chance is completed) then maximised interest lies in the game being re-played. The three losers want a rematch and only the winner wants things to remain as they are. There is no escape from this repetition as following the completion of every game interests will be maximised (three losers versus the lone winner) by replaying the game of chance.

Virtue theorists would seem no better placed. Given that there is no way to resolve the dilemma (within virtue theory itself) the technicians, as virtue theorists, might turn to drawing straws as in the previous examples. The difference with virtue theory is that it seems the losers ought to honor the result of

the game and let the winner of the lottery use the antidote, while the losers wait to die. I am somewhat dubious of a moral theory that both recommends and justifies the agent dying, although others seem to accept that morality may call upon the agent to embrace serious self-sacrifice. For example, Hursthouse claims, “here is an occasion where, say, if I speak out *as I should*, I am going to be shut in an asylum and subjected to enforced drugging”³³ (1999, p. 171). Rights theory fares much the same as virtue theory and offers no actual solution, only recommendations that the agent ought to accept the outcome, once a decision has been made. Each of the four agents has a right to life, but this in no way helps determine who gets the antidote. Rights-based theorists might also agree to the drawing of straws and might even agree that the winner had the right to the antidote but again it seems somewhat absurd to say that after the competition the losers should simply wait to die. Regardless of whether moral theory requires of an agent to lay down their life, my point is that none of the theories offer a solution to the actual dilemma. Whichever theory they subscribe to the agents might agree to a game of chance to decide who gets the antidote. However, there is no moral compulsion toward such an action and even should the agent agree to that game of chance it is the lottery and not the moral theory that determines the correct action. Further, upon completion of a game of chance, if the agents are consequentialists the losers will find themselves under no obligation to accept the result. For both egoists and utilitarians if the losers ask the question ‘what ought I do now?’ the original dilemma is recreated. Virtue and rights based theories, while offering no solution to the original dilemma, avoid the return to a state of conflict by advocating the honoring of the result in that the losing agents are morally required to die.

I contend that it is not a problem that egoism actively promotes conflict in the context of extreme scarcity (where the absence of conflict simply means accepting death) just as in certain contexts some moral theorists make room for the concept of a ‘just war’ (and in doing so implicitly accept that morality cannot always solve conflicts of interest). Egoism correctly determines that if life is our ultimate value

³³ My italics.

then in situations of extreme scarcity the resolution of the moral dilemma necessarily involves conflict.

The rules of *Jus ad Bellum* outline the generally accepted conditions of a just war. I would argue that, while individuals rather than states are involved, similar conditions occur where extreme scarcity is involved. The five conditions of *Jus ad Bellum* include that there be *just cause* for the conflict, that those involved have the *right intentions*, that there is a *public declaration* of the intent for conflict, that the conflict is a *last resort* and that there is *reasonable probability of a successful outcome* from the conflict. Through applying these rules to the technicians' scenario it can be clearly demonstrated that the conflict promoted by egoism meets the requirements of *Jus ad Bellum*. Firstly, each of the technicians has *just cause* given that the action is promoted by a well-developed theory of action and that the conflict is initiated only to protect an agent's life (where no alternative exists). Secondly, the technicians act for the *right intentions* given that the motivation to act is to protect the agent's own life. Thirdly, while it is unlikely the technicians would make a formal declaration to commence conflict, such a *public declaration* of intent is implicitly made. While individuals differ from states, and formal declarations of conflict are absent, given that all the involved agents recognise that a conflict of some form is inevitable then the declaration for the necessity of conflict is implicitly public³⁴. Fourthly, the conflict is undertaken only as a *last resort*. While the technicians ought to wait, hoping perhaps that further antidotes might eventuate, ultimately the only solution to the dilemma will be a conflict over the antidote, the only other alternative being death. In such cases it would seem that survival of the fittest is both necessary and preferable as the strongest is likely to have a greater chance of making a full recovery. The conflict will have a reasonable *probability of success* as, short of all the technicians being so incapacitated that no-one obtains the antidote, the probability is that the strongest will survive. Further, casualties from the conflict are limited to those that would likely die anyway.

³⁴ That is a rational agent will realise that the dilemma has no solution unless three of the agents are prepared to accept the outcome of some game of chance. The recognition that diplomacy will not lead to a solution (and that action must be taken) is also the recognition that conflict is inevitable.

In most cases egoism can resolve conflicts of interest and in the majority of cases it will promote co-operation as being in an agent's own interest. Failing to resolve conflict in cases of extreme scarcity is not a problem for egoism for the very reason that in these cases conflict *is* the solution to the problem. Unlike other theories, egoism is very clear about when conflict is justified: in times of extreme scarcity, where co-operation is not an option and where the agent's life is at risk through inaction.

§2.6 Conclusion

The position of this chapter so far has been mostly negative. I have been critical of the notion that egoism is not a moral theory on the grounds that morality is by definition other-regarding. In contexts where only the individual is affected, questions over how he/she ought to act, seem intuitively to be moral questions. To rule out egoism by definition alone accomplishes little and the question 'is egoism a moral theory?' simply slides to 'if egoism isn't a moral theory then why be moral?'. Further, I argued that egoism could not be excluded on the grounds that it fails to meet the basic function of conflict resolution. As demonstrated, in most cases egoism does offer conflict resolution and when it does not it seems that (i) other theories are equally unable to solve the conflict, and (ii) the conflict is justified under the conditions of *Jus ad Bellum*.

What then do I consider to be the definition of a moral theory? As a background against which egoism can be judged, in the chapters that follow I will adopt a minimalist and functional definition of morality; that function simply being action guidance. Like Kalin (1995) I accept that 'What ought I do...?' *is* a moral question and that any theory that offers a "coherent answer to it deserves to be regarded as a moral theory" (Kalin, 1995, 89). It follows that egoism is quite clearly a moral theory provided that it is coherent. I take a coherent theory to be one that fulfils the conditions of adequacy, that is, (at the very least) the theory is noncontradictory, consistent, and universal. Even these minimal requirements are open to challenge, but some bedrock must be accepted if any progress is to be made³⁵.

³⁵ I do not however dismiss the challenges to these criteria out of hand and where required discussion of the adequacy of these criteria will occur.

I will now turn my attention to whether egoism is a coherent theory beginning with an examination of Moore's (1962) proof that egoism is internally contradictory and thus not a coherent moral theory.

Chapter 3

Egoism is Contradictory

“[n]o more complete and thorough refutation of any theory could be desired.”
(Moore, 1962, 99).

§3.0 Ultimate Rational Ends and the Universal Good

In *Principia Ethica*, Moore formulated an argument that he believed was a complete refutation of ethical egoism. Moore attempts to prove that the theory of ethical egoism is self-contradictory and can therefore be dismissed not just as a valid moral theory but as a valid theory of any form³⁶. I will begin this chapter with an examination of Moore’s criticism, which is specifically aimed at Sidgwick’s (1913, §1) claim that ‘Egoism is a rational (moral) doctrine’. I will then examine some of the responses to Moore’s argument, most notably those raised by Broad and Smith, before developing a position based on Rasmussen that is not susceptible to Moore’s criticism. In doing so I will begin to build the axioms that according to ethical egoism denote right action.

Moore (1962, 99) paraphrases Sidgwick as claiming that “the Egoist may avoid the proof of utilitarianism by declining to affirm, either implicitly or explicitly, that his own greatest happiness is not merely the ultimate rational end for himself, but a part of Universal Good.” The proof of utilitarianism, according to Sidgwick (1913, Preface to sixth edition), is that from a non-interested (i.e. a God’s eye) perspective it is “reasonable to prefer the greater good [the good of the majority] to a lesser good [the good of the individual]”. However, the egoist can deny that his individual happiness, his good, is just a part of universal happiness and further claim that “It cannot be proved that the difference between his own happiness [and everyone’s happiness] is not *for him* all important.” (Sidgwick, ch. IV, §1). According to Sidgwick, given that relative to the specific agent there is an important difference between his happiness and a general maximisation of happiness then it is both reasonable and rational for the individual to give preference to his own good.

³⁶ However, as I will note in §3.1, the assumption that self-contradiction is sufficient to dismiss a theory is not accepted without exception.

Sidgwick's claims can be summarised as:

1. The egoist can claim the egoist's ultimate rational end is 'Good'.
2. The egoist can deny that 'my ultimate rational end' is 'an ultimate rational end in itself' and therefore claim it is not good-in-itself.
3. In short, the egoist claims 'my ultimate rational end' is 'a good' not 'good in itself'.

The egoist can claim that 'my happiness is my ultimate rational end and that the ultimate rational end is a good' without being committed to claiming that happiness is the objective ultimate rational end and therefore that happiness is a good in itself wherever it occurs.

4. The above concepts of good are part of the rational moral doctrine of egoism.

In brief, Sidgwick is outlining a position whereby an ultimate rational end is relative to an agent and, given that an agent's ultimate rational end is an agent's good, that good is relative to an agent. Further, Sidgwick is claiming it is rational for an individual to preference his/her own good over a (utilitarian) maximised good. Moore (1962, p.99) attacks both Sidgwick's claims regarding good and his conclusion (that egoism is a rational doctrine) as "absurd". According to Moore, Sidgwick's thesis is flawed in that it is never clearly defined what is meant by an "ultimate rational end" and it is "the use of such undefined phrases which causes absurdities to be committed in philosophy". Moore attempts to refute Sidgwick's claims via an analysis of the phrase 'ultimate rational end' in order to show that egoism - the view that we each ought to maximise our own good - is self-contradictory.

Moore claims that if 'X' is my ultimate rational end then it must:

1. Be rationally recognisable, by me, as 'truly good', and
2. In being an ultimate rational end it must be 'truly good in itself'.

That is, an agent's ultimate rational end must be rationally recognisable by the agent as good in itself. Given that Moore claims that 'the good' is simple, universal, and indefinable, if the agent recognises an ultimate rational end as good in itself then it must also be conceded that the agent's ultimate rational end (a

good in itself) is part of the universal Good (Moore, 1962, p.100). From this Moore derives the premise:

P1) If X is an agent's ultimate rational end then X is Good in itself.

Moore does not apply these exacting philosophical standards to his own thesis and defines Sidgwick's undefined phrase by reference to what is, according to Moore, an indefinable concept ('good'). While claiming that Sidgwick's absurd conclusions are the result of undefined phrases (insisting that 'ultimate rational end' *is* definable) Moore claims that good, as he is using it, is indefinable. However, it is unnecessary to enter into an argument as to what is and what is not philosophically definable as it is not essential to accept Moore's claim that Good is a simple and indefinable concept. All that needs to be accepted is Moore's claim that an 'egoist's ultimate rational end' is a 'good in itself', whatever 'good' happens to be. If Moore is correct then this analysis already leads to serious problems for Sidgwick's egoism; It seems that by egoism the *only* good is the one the agent determines *and* that each agent determines what that good is. Thus, "an immense number of different things are, *each* of them, *the sole good.*" (Moore, 1962, p.101).

Moore then turns his attention to the qualifying phrase '*for himself*' and denies that this qualification can have *any* meaning. All Sidgwick's '*for himself*' does is to introduce a dilemma, "the egoist's happiness must *either* be good in itself, and so part of Universal Good, *or else* it cannot be good in itself at all" (Moore, 1962, p. 100). Moore's claim can be broken down as follows: the clause '*for himself*' implies '*not* for others' or, at the very least, '*not necessarily* for others'. However, if 'X' is my ultimate rational end then 'X' (by premise 1) is good in itself and therefore equally good for others. If 'X' is not a good in itself then it cannot be an 'ultimate rational end' for me or anyone and the egoist has no ultimate rational reason to pursue 'X'. If 'X' is a good in itself then everyone has a rational reason to, and for Moore ought to, pursue 'X'. The egoist ought to promote 'X' for himself and everyone else ought to promote 'X' for the egoist, just as the egoist ought to promote 'X' for all others.

The following appear to be Moore's primary principles:

P1) If X is an agent's ultimate rational end then X is good in itself.

P2) If X is good in itself then everyone has reason to promote X for themselves and everyone else.

Assuming that 'X' equates to the agent's greatest happiness, Moore (1962, p.101) claims that no possible analysis of the egoist's claim that "his own greatest happiness is the ultimate rational end for himself can escape the implication that his own happiness is absolutely good; and by saying that it is *the* ultimate rational end, he must mean that it is the only good thing – the whole of the Universal Good", which leads to the formal contradiction of egoism.

(by egoism) Each agent ought to pursue their own ultimate rational end exclusively.

(by Moore's principles P1 and P2) Everyone has reason to promote the egoist's ultimate rational end and the egoist has reason to promote the same ultimate rational end for others.

Therefore egoism results in a contradiction in that the theory requires that an egoist pursue mutually exclusive goals. Egoism requires the pursuing of only the agent's ultimate rational end and the pursuing of the ultimate rational ends of others. Building on Moore's argument we can determine that an egoist (Agent₁) is committed to:

1. Agent₁ regards Agent₁'s happiness as Agent₁'s ultimate rational end.
2. If 'Agent₁'s happiness' is Agent₁'s ultimate rational end then 'Agent₁'s happiness' is *the* ultimate rational end (good in itself).
3. Agent₁ ought to pursue Agent₁'s ultimate rational end (*Agent₁'s happiness*) exclusively.

Furthermore, that some other egoist (Agent₂) is equally committed to:

1. Agent₂ regards Agent₂'s happiness as Agent₂'s ultimate rational end.
2. If 'Agent₂'s happiness' is Agent₂'s ultimate rational end then 'Agent₂'s happiness' is *the* ultimate rational end (good in itself).
3. Agent₂ ought to pursue Agent₂'s ultimate rational end (*Agent₂'s happiness*) exclusively.

If *the* ultimate rational end is ‘universal good in itself’ and therefore refers to the exact same thing in both cases, as Moore claims it must if it is *the* ultimate rational end, then under Moore’s analysis of obligation (namely that good in itself implies an equal obligation wherever it occurs and that we cannot make *ad hoc* restrictions of when the good obligates action) both agents have reason to act to promote the happiness of the other in direct contradiction with the principle of egoism.

If Agent₁ tries to deny this, claiming that only Agent₁’s own happiness is *the* ultimate rational end, then it follows that Agent₂ (another egoist) can make the exact same claim; only Agent₂’s personal happiness is *the* ultimate rational end.

However, this leads us to the position:

1. For Agent₁ only Agent₁’s happiness (X) is *the* ultimate rational end.
2. For Agent₂ only Agent₂’s happiness (Y) is *the* ultimate rational end.

From points 1 and 2 Moore would infer:

3. X is the ultimate rational end.
4. Y is the ultimate rational end.
5. $X \neq Y$ (Agent₁’s and Agent₂’s happiness are not the same)
6. If X then $\sim Y$ (If X is *the* (and therefore only) ultimate rational end and $X \neq Y$ then it follows Y is not *the* ultimate rational end).
7. $\sim Y$ is the ultimate rational end.

Ethical egoism leads to the logical contradiction Y is the ultimate rational end and not Y is the ultimate rational end. Therefore, by the principle of explosion (i.e., anything can be proved if the starting premise is a contradiction) if egoism is true then anything and everything is true.

If these arguments are sound and logical contradictions are necessarily false, then Sidgwick is mistaken, egoism is self-contradictory and not a rational doctrine. If ethical egoism is to meet even the minimal requirements of a moral theory then it must be shown where the presented arguments have gone amiss.

§3.1 Denying the importance of Logical Contradiction

To hold that a proposition is both true and a contradiction presents a major problem for classical logic; anything and everything can be derived (logically proven) if the initial premise is a contradiction.

1. $X \bullet \sim X$
2. X *Simplification of 1*
3. $X \vee (\text{Anything})$ *Addition*
4. $\sim X$ *Simplification of 1*
5. (Anything) *Disjunctive Syllogism 3,4*

Given the implications of accepting a contradiction as true it is usually assumed that showing a theory to be logically self-contradictory is a fatal criticism and sufficient reason to dismiss the theory as false. However, this is not necessarily the case and one valid response to the criticism that egoism is self-contradictory is simply to deny that the criticism of self-contradiction holds *any* compelling force. As Hindman (1997, para. 14) notes, for feminists “the charge ‘self-contradictory’ no longer necessarily disables another’s argument”. If classical logic is flawed then claims based on the logic are also flawed. While the criticism that a theory is contradictory may not be enough, in itself, to dismiss a theory it still seems far too important a consideration to overlook. At the very least, while resorting to the type of hierarchies abhorred by those who dismiss logical contradiction, it appears to be that a theory that isn’t contradictory is preferable to one that is. With this in mind I will move on to other methods of showing Moore’s criticism to be flawed that do not rely on the difficult task of tearing down classical logic.

§3.2 Broad’s response to Moore

Broad’s defence of Egoism against Moore is to attack the claim that X being Good in itself gives each of us reason to promote that good wherever it might appear. Broad (1952, 45) claims that if an agent is an ethical egoist then “he will assert that it is not his duty to produce good experiences and dispositions as such, without regard to the question of who will have them”. According to Broad, the egoist can deny Moore’s premise that ‘the good’ provides reason to act wherever it occurs. The egoist need not accept that even a ‘precisely similar good’ in others provides reason to act. According to Broad the egoist’s reasons to act are relative to who will benefit - and are compelling only when it is the egoist who will gain that benefit.

Moore's premise 2:

P2) If X is good in itself then everyone has reason to promote X for Y, themselves and everyone else.

According to Broad this ought to be formulated as:

P2) If X is good in itself and Y is an egoist then Y has reason to promote X *only* if Y will benefit from X

Broad's claim highlights a weakness in Moore's initial argument. The egoist need not accept that even a 'precisely similar good' for others (and it is not made clear exactly how Broad's 'precisely similar good' differs from Moore's 'good in itself'³⁷) provides reason to act and the egoist can *not* be accused of being contradictory in pursuing only his own good even if that good is precisely similar to the good of others.

Broad is successful in showing that egoism is not self-contradictory but I agree with Smith (2003, p. 582) that Broad is mistaken if he thinks this has "saved it simpliciter". If we allow Moore's first premise, and Broad has given us no reason to dismiss this, then an agent's ultimate rational end is good in itself. If something is good in itself and good in some way leads to obligations then "there can be no *ad hoc* restrictions on which of the things that are good create obligations" (Smith, 2003, p. 582). Given that the egoist need not accept that good in itself provides reason to act wherever it occurs, Moore's argument does not show egoism to be self-contradictory. However, while the contradiction has been removed the *ad hocery* remains. As Smith (2003, p.582) points out, if we allow Moore's first premise to stand unchallenged the egoist's ultimate rational end is a good in itself and if the egoist's 'ultimate rational end' is a good in itself then Broad's structuring of the egoist doctrine relies on an *ad hoc* condition for deciding when precisely similar goods obligate action³⁸. If Agent₁'s and Agent₂'s good is precisely similar and therefore, as Moore argues, equally obligatory then the egoist's claim that good (when I benefit) obligates action but a precisely similar good (when you benefit) doesn't, is *ad hoc*.

³⁷ Moore would likely argue that the reason why act 'A' and act 'B' appear to be precisely similar goods is simply that they are both parts of Good in itself. It is not that giving food to some starving person X and giving food to some starving person Y are cases of precisely similar goods but that giving food to the starving is Good in itself.

³⁸ Due to the 'good' in 'good in itself' being precisely similar wherever it occurs.

§3.3 Smith's Response to Moore

Smith correctly points out that Moore's argument is not just a dismissal of ethical egoism. Moore's argument can be generalised and in its generalised form it is dismissive of all ethical theories that hold that good is in some way relative.

When I talk of a thing as 'my own good' all that I can mean is that something which will be exclusively mine, as my own pleasure is mine ... is also *good absolutely*; or rather that my possession of it is *good absolutely*. The *good* of it can in no possible sense be 'private' or belong to me; any more than a thing can *exist* privately or *for* one person only (Moore, 1962, p. 99).

Moore's absolutist (agent-neutral) position on good means that commonsense morality, such as the position that I have a special ethical obligation to my family (relative altruism) is, like egoism, judged to be 'absurd'. We are equally obligated to act both for the good of those close to us and for total strangers, as we are equally obligated to act for the good of others and ourselves. "To the extent that we find this conclusion [that our obligation to promote the Good of family and strangers is equal] difficult to believe, the question we must ask ourselves is where [Moore's] argument goes wrong." (Smith, 2003, p.583).

In understanding where Moore's argument is flawed it is necessary to determine exactly what is entailed by the first premise and, more specifically, what Moore means by good. For Moore good in itself is simple, that is to say it is not a complex concept and cannot be defined by reference to its parts. Further, good in itself is non-natural; a simple property that is not identical with any natural property. For example, good cannot be defined as happiness. To do so is to commit the naturalistic fallacy. That is, it is not obviously apparent that the question 'is happiness in itself always good?' is simply a tautology (is happiness in itself always happiness?). Rather it would seem that the question is significant and that the answer is far from obviously yes.

Further, according to Moore, good in itself is an objective (non-relational) simple non-natural property. If X's ultimate rational end is good it is because the ultimate rational end itself has the property of goodness, it is not reliant on any relationship between the agent (X) and the ultimate rational end. However, the egoist need not

accept that good must be non-relational and can instead posit that the egoist's ultimate rational end is best defined in terms of Good_x ; that is, good is relative to a specific agent X. From this we can determine that an egoist holds all of the following:

- Agent₁'s ultimate rational end is $\text{Good}_{\text{Agent1}}$
- Agent₂'s ultimate rational end is $\text{Good}_{\text{Agent2}}$
- Agent₂'s ultimate rational end is not $\text{Good}_{\text{Agent1}}$
- Agent₁'s ultimate rational end is not $\text{Good}_{\text{Agent2}}$

This allows Smith to challenge Moore's first premise (if X is agent Y's ultimate rational end then X is good in itself), and deny that something being an agent's ultimate rational end leads to the conclusion that it must be good in itself. "In other words, what would follow from the fact that being F is good for A is not, as Moore would have it, that A's possession of the property of being F is good in an unqualified sense but, rather, that A's possession of the property of being F is Good_A " (Smith, 2003, p. 584).

Thus, retaining Broad's premise 2, only actions that affect an agent making the decisions obligate that agent to act, and modifying Moore's first premise to incorporate good as a relational concept, the egoist position becomes the noncontradictory position that:

- P1) If X is an agent Y's ultimate rational end then X is Good_y
- P2) If X is Good_y and Y is an ethical egoist then Y has reason to X

But, this is still insufficient either as a final defence of egoism or to entirely defeat the spirit of Moore's argument that the recognition of good, wherever it occurs, obligates action. Smith recognises that his defence is open to challenge and considers several counters to his arguments, then (routinely) rejects them all. While Smith solidly defends his thesis from most of the considered attacks, his rejection of the strongest counter-argument, that "the posited connection between goodness and obligation is *ad hoc*" (Smith, 2003, 594-598), is insufficient³⁹.

³⁹ I will not outline each of the individual counter-arguments here and will instead focus only on the strongest of the counter-arguments.

Working with Moore's concept of a simple, indefinable Good the goodness/obligation counterargument is formulated as follows:

Let us accept the assumption that good is (while simple and indefinable), as a matter of fact, relative to an agent.

Let us further assume that the act ϕ will bring about Good_{me}, while a precisely similar act ψ will bring about Good_{you}.⁴⁰

The egoist claims that I am obligated (or have good reason) to ϕ but I am under no obligation (or have no good reason) to ψ , but this is just an *ad hoc* condition on which goods obligate us to action. Good may be determined relative to an agent but once that good *is* determined, if that good is from a precisely similar act, then that good is in itself qualitatively equal. That is, the 'good' in the Good_{me} and the 'good' in the Good_{you} are equally good even if the determination of X being good is relative to an agent. Smith's Good_n shows only that the identification of good is relative to an agent, not that it differs in kind. If Moore is correct about the nature of obligations, then Good_{me} and Good_{you} once recognised, confer an equal obligation and to deny this is to place an *ad hoc* restriction on which goods obligate action.

Smith presents a defence to this criticism based on a psychological account of the desires of an agent. The rejection of this counterargument relies on the claims that:

1) Agent₁ believes ϕ is Good_{Agent1} (good for Agent₁) because Agent₁ believes Agent₁ would desire ϕ in a given situation if Agent₁ "had a maximally informed and coherent and unified desire set" (Smith, 2003, p. 597).

Thus, by universalisation:

2) Agent₁ also believes ϕ is Good_{Agent2} in a similar situation because Agent₁ believes Agent₂ would desire ϕ in a given situation if Agent₂ "had a maximally informed and coherent and unified desire set".

Smith argues that the most coherent pair of psychological states for Agent₁, with regard to position 1, is that Agent₁ both believes that in a certain situation Agent₁ would desire ϕ and actually desires ϕ . However, with regard to position 2, the

⁴⁰ Here I mean something like my obtaining food when I am starving versus you obtaining food when you are starving.

claim that ‘Agent₁ both believes that in a certain situation Agent₂ would desire ϕ and that Agent₁ actually desires Agent₂ ϕ ’ is no more coherent than ‘Agent₁ both believes that in a certain situation Agent₂ would desire ϕ and that Agent₁ is indifferent (or averse) to Agent₂ ϕ ing’.

Smith’s (2003, p. 597) defence is based on the claim that there is plainly “incoherence involved in my having beliefs about things that are Goods_{me} and yet having no desire that those Goods_{me} obtain” and no such “incoherence at all involved in my having beliefs about things that are Goods_{other people}, and yet having no desire at all that those Goods_{other people} obtain”.

This defence fails for several reasons. Firstly, unless some notion of second order desires are built into the equation we can quite clearly have beliefs about Goods_{me} yet not desire Goods_{me}. For example, someone may believe that they ought not smoke yet desire to smoke. This can be overcome by talk of second order desires; my having beliefs about things that are Goods_{me} and yet having no (second order) desire that those Goods_{me} obtain, even though I may have (addictive/weakness of will) desires to the contrary, is incoherent. However, there is no incoherence in holding the belief that I ought to give up X, and desiring to be in the position where I had given up X, but at this present point in time desiring X and the short term effects (pleasures) that come from X.

Secondly, if it is the case that any beliefs about Goods_{me} necessarily lead to a desire that Goods_{me} obtain then Smith must also hold that it is coherent to hold inconsistent desires and to embrace contradictory goals. It is possible that I hold the following beliefs about Goods_{me}, (i) that it would be Good_{me} if I gave up smoking (because I believe that smoking will cause health problems) and, (ii) that it would be Good_{me} to smoke (as I enjoy the pleasure of smoking). According to Smith if I believe *i* and *ii* represent Goods_{me} then I desire that both *i* and *ii* obtain. While it is logically possible, if inconsistent, to hold both of these desires it results in a contradiction to hold the desire that both *i* and *ii* obtain (that is, while I can hold inconsistent desires a contradiction occurs in desiring to bring about a state

of affairs that satisfies both desires). It cannot be the case that I necessarily desire that all beliefs I have about Goods_{me} come about.

Finally, it is unclear, even if it can be established that I must desire that those Goods_{me} obtain, how my *possibly* lacking a desire that a recognised Good_{other people} obtain for those other people frees me from an obligation to promote those Goods_{other people}. If the good involved is qualitatively equal, and it is the good that obligates my action, not my desires, then it makes no actual difference whether or not I desire that Goods_{other people} obtain. My desire to A and lack of desire to B may well explain why I act in manner A rather than B but it does not remove my equal obligation toward both A and B. If an egoist is not obligated to act in manner B it must be because the egoist does not recognise that B is necessarily a Good_{other people}, not that the egoist recognises but does not desire that Good_{other people} comes about. The reference to desires is an unsustainable *ad hoc* condition with regard to when Good obligates action.

§3.4 Rasmussen and Relative Values

Rasmussen (1999) considers a similar line of argument to Smith and while not posed as a refutation of Moore, and phrased in terms of human flourishing rather than the good, provides a distinction that allows us to draw the conclusion that in most cases the egoist cannot be obligated to act for the Goods_{other-people} for the simple reason that, in most cases, the egoist can never know what those Goods_{other-people} are.

Rasmussen denies that speaking of human flourishing as the ultimate rational end of human conduct commits us to agent-neutrality. According to Rasmussen (1999 p. 22) it is possible for an agent's flourishing to be both agent relative (Good_x) and universalisable (Good_x-in-itself). Rasmussen (1999, p.4) explicates the claim that human flourishing is agent relative as follows:

Human flourishing, G, for a person, P, is agent-relative if and only if its distinctive presence in world W₁ is a basis for P ranking W₁ over W₂, even though G may not be the basis for any other person's ranking W₁ over W₂. There is no human flourishing period, human flourishing is always and necessarily the good for some person or other.

Human flourishing is not an objective concept; what it is for a specific person to flourish is relative to a specific agent and, even more so, relative to that agent's worldview. Rasmussen (1999, p. 6) is claiming that no two specific cases of human flourishing can be the same, nor are they interchangeable. What constitutes a particular agent's flourishing is dependent on what it is for that agent to flourish given that agent's worldview.

Within Agent₁'s rational worldview (W₁), Agent₁'s *flourishing* (F₁) (whatever that might consist of) gives Agent₁ reason to act.

Within Agent₂'s rational worldview (W₂), Agent₂'s *flourishing* (F₂) gives Agent₂ reason to act.

and by universalisation:

Within any Agent_n's *rational* worldview (W_n), Agent_n's *flourishing* (F_n) gives a person (Agent_n) reason to act, and determines 'good' (where flourishing is an Agent's good) relative to the Agent_n

If attainment of *human flourishing* (whatever that might consist of) gives 'reason to act', then Agent₁ and Agent₂ must acknowledge the other's legitimate reason to act. However, F₂ does not obligate Agent₁ to act, nor does F₁ obligate Agent₂, for the simple reason that Agent₁ does not have direct access to W₂ (Agent₂'s worldview) and so cannot objectively determine what would constitute flourishing relative to Agent₂. *My flourishing* is my ultimate rational end and therefore, according to Moore, good in itself. However, it is only recognisable as good in itself within a specific worldview and that is both relative and unique to an agent. Thus good in itself is actually Good_{W(x)} in itself, that is, Good in itself relative to an agent (X) and that agent's worldview (W).

Rasmussen's notion of a worldview bears some resemblance to the Instrumental Theory of Rationality. According to Instrumental Theory reasons are inherently relative to some presupposed framework and while an agent can, and indeed must, reason within a framework agents cannot reason between these frameworks⁴¹. As Shaver (2002) notes all moral theories can be labelled instrumental, and therefore, in a trivial sense, all moral theories can be viewed as frameworks. 'The Good' or

⁴¹ As Foot (1997, p.315) notes these reasons can be moral, immoral or non-moral.

'Flourishing', however those concepts might be defined, are the goals of a moral agent and moral rules are simply instrumental in attaining that end. On a more serious level frameworks are not independent of a worldview and it is not possible for an agent to have a framework or theory unless they already hold to some worldview. Given that what leads to an agent's flourishing is relative to a worldview, not to some framework that may be layered atop that worldview, if ethical egoism is an instrumentalist theory it is so only in the trivial sense that a utilitarian giving to charity, as that maximises general happiness, is instrumentalist (Shaver 2002). An egoist may adopt several frameworks and reason within those frameworks, for example an egoist may adopt the framework of etiquette while at a dinner party and reason within this framework but what it is for the egoist to flourish is relative to the egoist's worldview not the temporarily adopted framework (even if the reason for temporarily adopting such a framework is the agent's flourishing).

Adding Rasmussen's concept that an agent's ultimate rational end (Moore's good in itself) is relative not just to an agent but within an agent's worldview we can reformulate the egoist doctrine as:

- P1) If X is an agent Y's ultimate rational end then X is Good_{W(y)} in itself.
- P2) If X is Good_{W(y)} in itself and Y is an ethical egoist then Y has reason to X.

That is, X is good in itself relative to agent Y and within Y's specific worldview. This position overcomes the goodness/obligation counter-argument to which Smith's position is vulnerable. While the act ϕ will bring about Goods_{W(me)} and the act ψ will bring about Goods_{W(you)} the good in each case, once determined, is qualitatively equal. It is not *ad hoc* to add that you are only obligated to act for the good only when that good is recognisable as such by you.

3.5 Moral Relativism

Harman (1983) considers flourishing as a basis for ethics but is critical of the merits of such an approach on the grounds that it leads to moral relativism. His criticism, while not a direct criticism of the position I have laid out so far, is important in that it may appear that defining obligation in terms of Good_{W(y)} must result in some form of moral relativism.

Harman (1983, 312-313) argues that flourishing “seems inevitably relative to one or another set of values. People with different values have different conceptions of ‘flourishing’, of the ‘good life’”. Further, “it is difficult to see how one rather than another conception of flourishing is to be validated simply in ‘the nature of things’ or in the ‘nature’ of persons - except in the sense in which different sets of values yield different conceptions of nature or of the nature of a person!”. Harman (1983, 320) goes on to claim that under such a form of moral relativism “there are no basic demands everyone has sufficient reason to accept as providing reason for action”.

At face value Harman appears to have made a valid criticism. Under the egoistic theory outlined Agent₁ may find the eating of other sentient creatures intolerable and detrimental to any form of flourishing while Agent₂ considers that fine meats are an essential part of the good life. However, if Harman believes that the essence of a good moral theory is its ability to provide an objective answer to such questions as the acceptability of eating meat then he sets a standard that is near impossible. Rather, I contend that the essence of a good moral theory is that it explains the nature of these differences, helps us to understand the implications of these differences and provides an objective answer as to how *I* ought to act (given my worldview) - not what is *the* objectively correct way to act. While this is a form of relativism, in as much as how I ought to act is determined by, and relative to, my worldview, it is a limited form of relativism. An agent cannot formulate whichever worldview the agent wishes and the boundaries which limit worldviews also limit what constitutes flourishing within those worldviews. If by relativism all that is meant is a theory that holds variable X is relative to some other variable then I have no quarrel with this label being applied here. However, as Machan (1985, 388) notes, “Harman’s talk about the alleged relativism of any flourishing ethics really amounts not to relativism but to contextualism.” It is not relativism (as the term is commonly used) to note that what constitutes flourishing depends on the context, the rational worldview that a specific agent holds, anymore than it is “relativism in medicine to recognise that general principles of human health apply differently to different people” (Machan, 1985, p.389).

Consider the following cases:

Within Agent₁'s worldview animals are, as sentient creatures, members of the moral community and to eat a fellow member of the moral community is, to Agent₁, a horrendous thought. Needless to say, committing horrendous acts is detrimental to Agent₁'s flourishing and therefore Agent₁ will claim such acts *are* wrong - although they are actually wrong in themselves only with reference to Agent₁ and Agent₁'s worldview.

For Agent₂ animals are a living resource and while Agent₂ may believe animals ought to be treated humanely (for anthropocentric reasons) considers there is nothing wrong with raising them for food. The enjoyment Agent₂ gets from eating meat enhances Agent₂'s life and leads to Agent₂'s flourishing and is judged by Agent₂ to be good - however, it is only actually good relative to Agent₂ and Agent₂'s worldview.

The moral judgments themselves are incommensurable without an understanding of the worldviews the judgements are relative to and while we cannot have direct access to the worldviews of either Agent₁ or Agent₂ what we do know is that what constitutes flourishing for Agent₁ and Agent₂ is dependent upon those worldviews. Once it is recognised that the source of the differing moral conclusions is due to differences in the agents' worldviews (with regard to the moral status of animals) it may be possible to objectively discuss why these conclusions were reached.

In ruling out all forms of relativism Harman sets the moral bar too high. It also seems that his critique contains several misunderstandings. Firstly, in claiming that flourishing is relative "to one or another set of values" Harman appears mistaken. Flourishing is not relative to some selected set of values, rather it is objectively determined within an agent's worldview⁴². The theories Harman is objecting to use flourishing as 'the standard', as 'the value' for making moral judgements, and this being the case, as Machan (1985, 388) notes, "[f]lourishing, then, cannot be relative to one or another set of values since there are no values apart from flourishing, not at least in the morally relevant sense".

Secondly, it seems unlikely that Harman is correct in saying that a position (which incorporates a limited relativism) leads to there being 'no basic demands everyone has sufficient reason to accept as providing reason for action'. Given the

⁴² Exactly how flourishing is determined relative to a worldview will be discussed in Part 2. In brief the worldview we hold determines what it is to flourish in the world.

constraints on the worldview we can rationally construct there may well be areas of overlap and, even given that what constitutes flourishing for each individual agent is determined from within that agent's worldview, it is possible (and I will suggest probable) that individual agents will find they have sufficient reason to (and are perhaps obligated to) carry out actions that are, in application, precisely similar. Further, as human knowledge advances the constraints on worldviews that can be rationally constructed will likely become ever tighter. As this happens the overlap between worldviews, and the moral obligations that arise from those worldviews, will become ever more similar.⁴³

Even where worldviews lead to incommensurable moral conclusions, such as Agent₁ determining it is immoral to eat meat and Agent₂ determining the same act is morally acceptable, an understanding of the differing conclusions can be reached by an examination of the worldviews on which they are based. If either agent can be shown that their worldview is flawed, or that a competing worldview is superior, then what constitutes flourishing for that agent may well change and with it what constitutes correct action. For example, when worldviews contained the assumption that Africans were less than human it was a simple matter to formulate the moral conclusion that using them as slaves was morally acceptable. As science progressed, and human knowledge increased, it became ever harder to rationally hold on to such a view, and as that worldview became far harder to justify so did the notion that slavery was morally acceptable.

3.6 Moorean Challenges and Naturalistic Fallacies.

Returning to Moore, I will consider one final point before dismissing the criticism that egoism is internally contradictory. It might be argued that I have moved, somewhat loosely, between talk of 'the good' and 'flourishing' throughout this chapter and in doing so have conflated the two and committed the naturalistic fallacy. Firstly, I do not believe this to be the case. In defending egoism against charges of contradiction I have been developing a standard for value judgements and in this respect the substitution of 'flourishing' for 'the good' was well motivated and necessary. Secondly, I do not believe that Moore's open question

⁴³ The defence against Harman's criticism outlined in this section has been very brief. The reason for keeping it brief is that the defence depends upon a deeper analysis of worldviews; I will undertake such an analysis in Chapter 6 when considering another of Harman's criticisms.

‘Is flourishing itself Good?’ or ‘Ought we flourish, rather than do something else?’ makes any sense. In postulating flourishing as a standard by which right action can be judged, to ask if this standard is itself good is “presumptuous since it treats the question at hand as if it already had a good solution.” (Machan, 1985, p.389) ‘Is flourishing itself good?’ can only be answered if some independent standard for good already exists. While the question appears at face value to be important it either:

(i) Reduces to a tautology; ‘Is flourishing flourishing?’

If flourishing is the standard for good then it makes no sense whatsoever to ask if it is good. If, as will be argued throughout this thesis, flourishing is the standard by which good is to be measured then to ask ‘is flourishing good’ is simply not to ask a question. All an inquirer asks in questioning if flourishing is good is ‘if good is good’ and that is not a question.

or,

(ii) Requires a reference to some independent pre-existing standard of the good.

That is, if we cannot use flourishing as our standard for good then the question cannot be answered unless we have some *other* standard reference for good. However, according to Moore’s naturalistic fallacy we can have no such standard.

In this case the question ‘is flourishing good?’ is really ‘is flourishing good by reference to some indefinable concept?’ a question that is unanswerable and therefore meaningless.

Moore’s question, as phrased, is either meaningless or unanswerable. To make use of an analogy, the following question might be posed, ‘Is X really light?’ That is, “when a theory is advanced as to what light is, people can ask ‘but is that really light?’ as if they already know what light is.” (Machan, 1985, p.389). To ask if X is really Y (where X is an explanation of Y) is meaningless in that the question is unanswerable except by reference to X or to some other theory; a theory that would then itself face the exact same question.

We should not ask if X is really Y but rather: Is X the best theoretical explanation of Y that is available to us? Similarly, the real question is not: ‘Is flourishing really good?’, but ‘Is flourishing the best explanation of good?’, i.e., ‘is flourishing the best candidate for serving as the standard for value judgements?’ – At this point the answer to the latter question appears to be yes.

3.7 Conclusion.

The purpose of defending egoism from the criticisms in this section has been twofold. Firstly, if egoism was to fall to these criticisms there would be no point continuing with this line of investigation and thus a response to the criticisms is essential. In this respect I believe I have successfully shown that egoism, correctly formulated, involves no logical contradiction with regard the egoist's concept of 'the Good'.

The second purpose, as intimated above, was to determine how to best formulate the theory of egoism. In avoiding the criticism of contradiction it has been determined that the egoist must hold to a version of the good that is relative to a specific agent's worldview.

To briefly recap, in Chapter 2 I argued that egoism provided an answer to the question "what ought I to do?" and that any theory that provided an answer to this moral question qualified as a moral theory. In this chapter I have argued that egoism is not contradictory and made a first attempt at developing the initial premise of egoism.

If Φ is $\text{Good}_{w(y)}$ and Y is an egoist then Y has reason to bring about Φ '.

If an agent recognises some 'state of affairs' as good, where good is relative to an agent's worldview, then that agent (as an egoist) has strong moral reason to act in a manner that will bring about that 'state of affairs'. I have not claimed that ethical egoism provides overriding reasons to act; but have instead argued that the very concept of overriding reasons for action is logically incoherent. An egoist can, and should, accept that there can be other reasons to act. In deciding how to act a process based on legal, financial, aesthetic, or religious theory could be applied and reasons for acting in a specific manner determined. An egoist must also accept that there are other moral theories and application of any of these theories will also provide a reasoned answer to the question how should I act.⁴⁴ In becoming an ethical egoist what an agent does is not to deny that these other reasons for action exist but to give preference to self-interested reasons for action, claiming to be morally justified in doing so. Consider the task of buying a new

⁴⁴ I am assuming that all these theories meet the minimum requirement of adequacy.

fridge. An agent may simply apply financial reasoning; purchasing the fridge that best meets the agent's budgetary restrictions. Likewise, the agent may apply only aesthetic reasons throwing financial concerns to the wind and purchasing the fridge that most matches the agent's aesthetic sensibilities. However, if the agent asks 'which fridge ought I buy?' then the agent will turn to moral theory and the moral implications of such a purchase. Which moral theory an agent will turn to depends on which theory the agent believes offers the best explanation, or theory, of how an agent ought to act.⁴⁵

To allow that an agent has 'strong reason to bring about Φ ' rather than 'an agent is obligated to bring about Φ ' enables us to avoid falling into logical paradoxes, without rendering ethical statements meaningless. If, when faced with a tragic dilemma such that an agent has equal prima facie duties to act in mutually exclusive ways then a strong sense of 'ought' can result in logical contradictions. To say that 'an agent is obligated to bring about Φ , that is that an agent must act to bring about Φ , but that the agent is not going to act in such a manner' lacks sense. As Cordner (2001, p. 542) notes, if an agent claims 'I have to do X but I'm not going to do X' then at the point the agent commits to 'I'm not going to do X' the agent has already dismissed both the force of the obligation and the belief that the agent has to do X. However, if the moral 'ought to' means only that an agent has 'strong reason to' then this sort of contradiction is overcome.

The egoist doctrine that 'an agent (Y) on recognition of some $\text{Good}_{w(y)}$ ought to act so as to bring about that good (Φ)' means only that the agent (Y) has 'strong (moral) but non-obligatory reason to act so as to bring about that good'. An agent can believe that 'I ought to act to bring about Φ ' and continue to claim 'yes, I ought to act to bring about Φ ' even as the agent commits to acting in some other manner. 'Yes, I have strong reason to act to bring about Φ but (for some other reason) I am not going to' expresses that an agent understands the reasons for acting to bring about Φ but on this occasion is determined to act otherwise, that is the agent has decided to give preference to some other reason for acting. However, while such a position avoids contradiction it may lead to inconsistencies or incoherence. I will now my attention to such criticisms.

⁴⁵ The position I am developing is subjective in the limited sense that, as I will go on to argue, it is logically impossible to claim one coherent moral theory is superior to another.

Chapter 4

Inconsistency & Incoherence

§4.0 Introduction

Ethical egoism provides an agent with strong reason to carry out an action rather than an obligation to act. However, if an agent (y) recognises that Φ is $\text{Good}_{w(y)}$ and therefore has strong reasons for acting so as to bring Φ about, then it is also reasonable to assume that the agent desires that Φ come about.

Accepting that tragic dilemmas exist, situations will arise where an agent would desire (if it were possible) to take mutually exclusive actions. This being the case it seems likely that all moral theories will, on occasion, lead to inconsistent desires. However, it may simply be that ‘given the way the world is’, tragic dilemmas are unavoidable and the fact that moral theories may, on occasion, result in an agent formulating inconsistent ideas does not, in itself, present a major problem. Perhaps we need simply accept that situations arise for which no perfect solution exists.

However, the question ‘what ought I do?’ does not surface only when an agent is left facing a tragic dilemma. If the consequence of attempting to answer this question by evoking a specific moral theory is that on each and every occasion the agent formulates inconsistent desires then it is unlikely any maximally informed agent would give preference to this kind of theory. It is the claim that egoism is such a theory that underlies Medlin’s (1957) criticism. Medlin argues that attempting to universalise egoism leads to a general inconsistency to the point where the theory becomes incoherent and offers no moral guidance. If this criticism is successful then egoism is not a moral theory (and perhaps not a valid theory at all) in that it fails to meet even the minimalist definition that a moral theory provides an answer to the question “what ought I to do?”

§4.1 Ultimate Principles and Ethical Egoism

Medlin claims that ultimate ethical principles are necessarily boldly asserted and are not the result of philosophical arguments. “To arrive at a conclusion in ethics one must have at least one ethical premiss. This premiss, if it be in turn a

conclusion, must be the conclusion of an argument containing at least one ethical premise. And so we can go back indefinitely but not for ever.” (Medlin, 1957, p. 111). Medlin’s claim is that at some point all philosophical arguments either:

- (i) become circular and rely on the conclusion of some later argument as an essential premise.
- (ii) assert some naturalist claim and attempt to derive an ‘ought’ from an ‘is’, or
- (iii) rely on the claim that bedrock has been reached and assert a value premise as being foundational.

Arguments that rely on (i) or (ii) are usually dismissed on the grounds that (i) circular argument are fallacious⁴⁶, and (ii) the claim that it is not possible to derive a value (how we ought to act) from a fact (how the world is)⁴⁷. This leaves only (iii) arguments that make a claim that some principle is bedrock and must be accepted, without argument, if *any* progress is to be made. In the case of moral theory this bedrock usually takes the form of an ultimate ethical principle. If this indeed is the case, then as Medlin notes, this principle ought to be closely scrutinised for “there is no room for reason even to go wrong”. (1957, p.111) Any principle that is to stand, or be preferenced, as an ultimate principle of moral value must undergo the most rigorous examination. Medlin undertakes such an examination of the ultimate principle of egoism and claims that if egoism is a description of how we ought to act then its principle, which is nothing more than an “expression of incompatible desires” (Medlin, 1957, p.118) is unsuitable as a candidate for an ultimate moral principle. Further, if the ultimate principle on which egoism is based is dismissed then the theory itself must also be dismissed.

Medlin begins by setting up the basic criteria for something to be classified as a moral theory. Firstly, Medlin (1957, pp. 112-123) claims that ethical egoism, if it is a moral theory, must be universal⁴⁸. If an agent subscribes to a doctrine that cannot be promulgated, a doctrine that must be held in secret, then according to Medlin that agent holds no moral doctrine at all. Medlin implicitly accepts the claim that moral theory must be public and that a theory, held in secret, thus

⁴⁶ Although some trivial circularity may be deemed acceptable.

⁴⁷ I will not engage with this claim here.

⁴⁸ For Medlin being universal appears to necessitate being public.

rendering an agent unable to enter moral debate or to give moral advice is something other than a moral theory. Medlin gives no argument as to why a moral theory must be universal or why it must be public⁴⁹. However, given that this very discussion is about how agents ought to act I will accept that the discussion is, and indeed must be, about a public moral theory. It has not been determined that ‘being universal’ is part of the minimum requirement of adequacy. However, for the sake of argument I will accept Medlin’s stipulation that a theory must be universal, at least to the extent that any agent in the same circumstances facing the same moral choice ought, in applying a specific theory, reach the same moral conclusion.

Secondly, Medlin (1957, p. 113-114) claims that egoism must be formulated as a categorical theory, that is, it must be consistent with the claim that the egoist doctrine incorporates an ultimate moral principle and that we ought to act in accordance with that principle “*because that is what we ought to do*”. If the egoist tries to avoid this formulation and instead claims that ‘an agent ought to act in accordance with the principles of egoism because...’ then the egoist is not an egoist. The ultimate principle subscribed to is whatever follows the ‘because’, and given the denial of the categorical formulation the principle must be something other than egoist in nature.

To summarise, Medlin claims that:

- (i) A moral theory must be public and universal.
- (ii) A moral theory must be categorical not hypothetical.

§4.1.1 Medlin’s attack on Ethical Egoism

Medlin (1957, p.114) briefly attacks egoism on the grounds that, since it cannot be promulgated, egoism fails to meet the basic requirements of a moral theory, namely that such theories must be public. If an agent is an egoist and the primary

⁴⁹ Medlin (1957, p.114) does note that an agent could hold to a form of private egoism and while Medlin claims such a doctrine, if it is a doctrine at all, is not an ethical one he admits that an individual egoist may quarrel with this. Medlin offers only the following reply to someone who holds such a position “Let him call himself what he will, it makes no difference. I’m a philosopher, not a rat catcher, and I don’t see it as my job to dig vermin out of such burrows as individual egoism”. I contend that individual egoism is not as impregnable as Medlin fears but will delay discussing the concept of a private morality until §4.2 and §4.2.1

concern of an egoist is to further the egoist's own interests then, according to Medlin, in many cases to promulgate the egoist doctrine and to recommend that others give primacy to their own interests would be to frustrate the egoist's own interests and hence to act contrary to the principles of egoism.

With reference to the infected scientists scenario (§2.5), if one of the scientists was an egoist and the remainder subscribed to virtue theory then it would be contrary to the egoist's doctrine to promulgate the doctrine and in doing so risk persuading the others that egoism is a preferable ethical theory. Further, if after the drawing of straws the egoist is fortunate enough to be the winner then the egoist is better off applauding the others' virtue (as they wait to die), keeping his beliefs about how they ought to act to himself. I have already claimed that the infection case is a special one and bereft of easy moral answers. However, Medlin's claim is far more general, namely that in ordinary everyday cases it will be contrary to the interests of the egoist to promulgate his doctrine. In promoting their own interests, others will invariably have an impact on the egoist's interests and thus it is contrary to the egoist's interests to convince others to adopt this doctrine.

Medlin considers two responses the egoist may raise. Firstly, the egoist might claim that the doctrine can be promulgated as long as it is fully understood what it is to be in an 'agent's ultimate interest'. I have already advanced this line of argument in attempting to show that the egoist doctrine tends to promote co-operation and that the egoist has a self-interest in avoiding conflict (§2.5). Medlin thinks it will always be possible to force this type of defense into hypothetical egoism and ultimately show that some general principle other than egoism is being defended⁵⁰. Secondly, the egoist may claim that the doctrine may be promulgated as long as the agent promulgating the doctrine takes pleasure from doing so (that is, it is in that agent's self-interest to promulgate the doctrine). However, Medlin concludes, rightly I think, that this second defense is just a disguised form of individual egoism in which, just as a matter of chance, the egoist takes enjoyment from promulgating the doctrine.

⁵⁰ Claims such as we will all be better off if we act out of self-interest showing the assumed egoist to be a confused utilitarian.

The criticism that egoism cannot be promulgated is serious and I will return to this claim in later sections. However, I will set it aside for the moment, just as Medlin does (believing it unnecessary to develop this line of attack), in order to move on to a criticism that Medlin claims is “both fatal and irrefutable”.

§4.1.2 Ethical Egoism: Inconsistency and Incoherence

Medlin (1957, p. 166) claims that the egoist’s belief that ‘I should act for my self-interest’ can be equated with ‘I want to be happy’ and ‘I ought to do whatever makes me happy’. Further, Medlin insists that this also involves the attitude ‘I want myself not to care about others’. However, while it may be that the egoist has no primary reason to act for the happiness of others (unless doing so promotes the egoist’s own happiness) this does not mean that the egoist must refrain from doing so, nor does it entail that the egoist does not care about others. Nothing in the egoist doctrine suggests that he ought not, or does not, care about others only that the preferred reasons for action involve primarily self-interest. Perhaps Medlin means that the egoist does not care about others as an end in itself. Even so, the claim seems insufficient to support Medlin’s view.

The egoist doctrine developed so far (§3.7) consists of the conjunct:

I (Agent₁) in recognising Φ as $\text{Good}_{W(\text{Agent}_1)}$ (that is, Φ is good in itself relative to Agent₁’s worldview) have good reason to bring Φ about.

and

I (Agent₁) in acknowledging that Ψ is $\text{Good}_{W(\text{Agent } n)}$ (that is, Ψ is good for some other agent ‘n’ relative to that agent’s worldview) do not have good reason to bring Ψ about.

Inserting Medlin’s emotivist view (with regard to moral language) the claims that Φ is good relative to my worldview, is in my self-interest, and that I have good reason to Φ equates to me wanting and desiring Φ . Accepting that I do desire that Φ comes about the conjunct can be reformulated as:

I (Agent₁) in recognising Φ as $\text{Good}_{W(\text{Agent}_1)}$ (that is, Φ is good in itself relative to Agent₁’s worldview) desire that Φ comes about.

and

I (Agent₁) in acknowledging that Ψ is $\text{Good}_{W(\text{Agent } n)}$ (that is, Ψ is good for some other agent ‘n’ relative to that agent’s worldview) do not desire that Ψ comes about.

Medlin, applying the principle of universalisability, claims that the principle applied to Agent₁ must equally be applied to Tom (Agent₂), Dick (Agent₃), and Harry (Agent₄) and that this “can be expressed by an infinite number of avowals” (Medlin 1957, p. 115).

I (Agent₁) in recognising Good_{W(Agent1)} *want (desire)* Good_{W(Agent1)} to come about.

and

I (Agent₁) in recognising Good_{W(Agent2)} *want (desire)* Good_{W(Agent2)} to come about.

and

I (Agent₁) in recognising Good_{W(Agent3)} *want (desire)* Good_{W(Agent3)} to come about.

and

I (Agent₁) in recognising Good_{W(Agent4)} *want (desire)* Good_{W(Agent4)} to come about.

and with reference to the second clause of Medlin’s conjunct and in accordance with egoism.

I (Agent₁) in recognising Good_{W(Agentn)} *have no desire* that Good_{W(Agentn)} comes about.

I have already argued that an agent is not obligated to act for another’s good because good is relative to an agent’s worldview and in most cases Agent₁ will not have access to Agent₂’s worldview and thus no knowledge of Agent₂’s good. However, although an agent, in most cases, cannot know the specifics of another’s good the agent does know that there is something that represents good, albeit relative to that agent’s worldview. Further, Agent₁ must accept that the good of others (Good_{W(Agentn)}) may be identical to Agent₁’s own good (Good_{W(Agent1)}) or, that some other agent in pursuing what that agent recognises as good (Good_{W(Agentn)}) may act in a manner that is detrimental to Agent₁’s good.⁵¹ Consider some scenario where the good of several agents requires the mutually exclusive acts (such as the laboratory/antidote scenario in §2.5 where the good of one agent, i.e. getting the antidote, excludes others from attaining their good).

According to Medlin:

⁵¹ Medlin (1957, p.117) does consider the counter-arguments that (given the world was different than it is) everyone could obtain whatever was good relative to that agent. However, as Medlin goes on to note, this is simply not the case in this world and everyday examples of conflict are commonplace. Further, it is pointless for the egoist to claim ‘if the world were different...’ as to do so would be to slip into hypothetical egoism and ultimately to argue as a utilitarian; if the world was such that acting as an egoist maximised happiness (utilitarian ultimate principle) then we ought to act as egoists.

By universalising egoism: I desire that I get the antidote and, I desire that Agent₁ gets the antidote and, I desire that Agent₂ gets the antidote and, I desire that Agent₃ gets the antidote and, I desire that Agent₄ gets the antidote.

By egoism itself: I desire that I get the antidote and therefore also desire that others (Agents₂₋₄) do not get the antidote.

Clearly such desires are inconsistent and incompatible. Further, if these desires are equal in strength (and according to the principle of universability they must be equal) then the egoist theory provides no action guidance. Faced with equal, mutually exclusive, and competing desires, no answer to the question ‘what ought I to do?’ has been provided. This leads Medlin to conclude that egoism is not only inconsistent but also incoherent. In Medlin’s view egoism, taken at face value, only appears to be a theory of how to act; once egoism is universalised it provides no such direction. Rather, an agent is presented with an infinite number of conjuncts that each provide equally compelling force toward incompatible ends. If a theory of action fails to give any action guidance, or presents guidance that cannot be followed, then nothing has been said. Medlin concludes that universalised egoism fails to provide action guidance, is therefore incoherent and in failing to answer the question “what ought I to do?” is not even a moral theory.

§4.2 Kalin: In Defence of Egoism

Kalin (1995, p. 83) attacks the underlying logic of Medlin’s argument and claims an error has been made in universalising the egoist principle. Moving from the egoist principle that ‘I ought to (or have strong reason to)...’ to ‘others ought to (or have strong reason to)...’ does not result in the type of conjunct that Medlin envisages. An agent who subscribes to egoism is someone who desires to follow, or give preference to, the dictates of egoist doctrine:

(a) I (Agent₁) desire to act in accordance with the dictates of egoism.

Universalising this principle means that the egoist (Agent₁) believes others (Agent_n) ought to act in accordance with the dictates of egoism and, according to Medlin, wants others to act in accordance with egoism. The correct conjunct to (a) is:

(b) I (Agent₁) desire that others (Agent_n) act in accordance with the dictates of egoism.

If the clause ‘dictates of egoism’ is fully expanded in terms of the egoist ultimate principle (as developed so far), universalising egoism results in the following conjunct:

I (Agent₁) in recognising Φ as $\text{Good}_{W(\text{Agent}_1)}$ (that is, Φ is good in itself relative to Agent₁’s worldview) have good reason to bring Φ about.

In Medlin’s terms, Agent₁ wants, and desires that Φ come about.

And

I (Agent₁) want others (Agent_n) in recognising Ψ as $\text{Good}_{W(\text{Agent}_n)}$ to acknowledge they have good reason to bring Ψ about.

In Medlin’s terms, Agent₁ desires that Agent_n desires Ψ comes about.

Agent₁ does *not* want $\text{Good}_{W(\text{Agent}_n)}$ to come about. Agent₁ wants Agent_n to want $\text{Good}_{W(\text{Agent}_n)}$ to come about.

Kalin demonstrates the principle by analogy to sports and games. In any sports match a team does not want the opposition to win. Rather, the team wants the opposition to want to win (in order that good competition follows). But, at the same time as a team wants the opposition to want to win, they also want the actual result to be that the opposition loses.

The infinite conjunction of avowals⁵² that the egoist holds is:

I (Agent₁) in recognising $\text{Good}_{W(\text{Agent}_1)}$ *want (desire)* $\text{Good}_{W(\text{Agent}_1)}$ to come about.

and

I (Agent₁) in recognising $\text{Good}_{W(\text{Agent}_n)}$ *want (desire)* Agent_n to want $\text{Good}_{W(\text{Agent}_n)}$ to come about.

and

I (Agent₁) do *not want (desire)* $\text{Good}_{W(\text{Agent}_n)}$ to come about unless $\text{Good}_{W(\text{Agent}_n)}$ is beneficial ($\text{Good}_{W(\text{Agent}_1)}$) to Agent₁.

Kalin claims that such a universalised position is both consistent and coherent and, as Sidgwick notes, even though from the point of view of the universe, it may be preferable to act for a greater good than a lesser good it is “also undeniably reasonable for the individual to prefer his own” (Sidgwick, 1913, Preface to Sixth Edition, p. xviii).

⁵² Medlin declares the conjunct to be ‘infinite’. The conjunct refers to the conjunction of the desires of all humans (and possibly all future generations) even so the conjunction is really only very large not infinite.

Surprisingly, given that he is writing in defence of egoism, Kalin (1995, p. 86) agrees with Medlin that the egoist cannot promulgate the doctrine. Further, given that the doctrine cannot be promulgated, Kalin (1995, p. 86) claims the egoist is also prevented from joining moral discussion, giving or receiving moral advice, establishing laws based on egoism, teaching morality to others, or justifying actions to others. Kalin claims that the egoist's inability to promulgate the doctrine poses no significant problem for egoism because there is no reason why an egoist *must* promulgate the doctrine. According to Kalin (1995, p. 89) egoism is essentially a private morality.

§4.2.1 Kalin, Wittgenstein, and Private Morality

Assuming that the logic of Kalin's defence is sound, like Hauptli (2002) I am left wondering what exactly the purpose of a private morality is. Early in his thesis Kalin (1995, p. 83) claims that the egoist is not affirming a moral doctrine simply so he can get his goodies out of life (no moral theory being necessary for such a course of action). However, it appears that the only function performed by Kalin's version of egoism is the self-justification for an agent 'getting his goodies out of life'. Given that Kalin's egoist must hold his doctrine in secret and apply the theory without recourse to any external criterion it would appear that there is no way for an agent to distinguish between acting out of self-interest and acting out of desire that is perceived (by the agent) as self-interest.

I will argue that a strong parallel exists between attempting to hold a private theory and attempting to employ a private language and that Kalin's private morality is subject to Wittgenstein's criticism of private language.

Let us imagine the following case. I want to keep a diary about the recurrence of a certain sensation. To this end I associate it with the sign "S" and write this sign in a calendar for every day on which I have the sensation. - I will remark first of all that a definition of the sign cannot be formulated. - but still I give myself a kind of ostensive definition. How? Can I point to the sensation? Not in the ordinary sense. But I speak, or write the sign down, and at the same time I concentrate my attention on the sensation. But what is this ceremony for? for that is all it seems to be. A definition surely serves to establish the meaning of a sign. - Well that is

done precisely by the concentration of my attention; for in this way I impress upon myself the connection between the sign and the sensation. But "I impress it on myself" can only mean: this process brings it about that I remember the connection right in the future. But in the present case I have no criterion of correctness. One would like to say: Whatever is going to seem right to me is right, and that only means that here we can't talk about 'right'. (Wittgenstein, 2001, Part 1, §258)

What I take Wittgenstein to mean is that, given I have a private mental state to which I refer by a specific sign 'S', there is no criterion of correctness by which I can test if the sign 'S' is rightly applied to some later mental state. There is no possible test to confirm either that I am using 'S' correctly or whether the experienced mental state is similar, in all relevant respects, to the earlier one. Talk of my private mental state now being the same type of mental state 'S' is meaningless. The most that can be claimed is that the private mental state I am experiencing now *seems* like the same type of mental state I previously referred to as 'S'.

I suggest something similar happens when Kalin's egoist attempts to apply his private morality. The egoist, facing a moral choice, internally questions 'What ought I do now?' and claims I ought to act to bring about Φ because Φ is in my best interest. However, given that the reasons the egoist believes ' Φ ' to be in his best interest are private (and according to Kalin cannot be shared or discussed without violating the egoist principle) then there is no way of determining the truth of the claim that Φ is actually in the agent's best interest. Given the compelling force of human 'wants' and 'desires' it is easy to imagine that an agent could mislabel a 'want', rather than a 'best interest', as Φ . An agent wants to bring about Φ because that agent believes Φ to be in his best interests. However, the only test as to Φ actually being in the agent's best interest is internal, so there is no testable difference between what *seems* to be in the agent's best interest and what actually *is* in the agent's best interest.

I am not claiming that the argument against private morality is the same as the private language argument or that the argument can be directly carried across. However, the structure of the two arguments does suggest an interesting parallel. Further, if we adopt Kripke's (1982, pp. 109-110) view that the private language

argument is an example of the impossibility of following a private rule then the argument against private morality is another, strikingly similar, example. Kripke goes on to claim, the conclusion to the private language argument can be found prior to its specific formulation; “Hence it is not possible to obey a rule ‘privately’: otherwise thinking one was obeying a rule would be the same thing as obeying it.” (Wittgenstein, 2001, Part 1, §258) For Kalin’s private egoist there is no difference between following the dictates of egoism, and acting out of self-interest, and acting in a manner believed to be the dictates of egoism, but potentially nothing more than fulfilling some desire. It may well turn out that the egoist is acting out of self-interest, but without some means of verification, a verification that must be external to the agent, the claim that a moral doctrine *is* being adhered to must be abandoned. A minimum requirement of adequacy appears to be that for a theory to be applied that theory must be public.

§4.2.2 Kalin, Carlson, and What the Egoist Wants

Even assuming Kalin’s thesis can be reformulated to overcome the outlined problem in holding a private morality the thesis is still open to a second form of attack, namely that Kalin hasn’t actually universalised egoism but rather presented some type of sophisticated individual egoism.

Carlson (1973, p. 28) attacks Kalin’s notion of (weak) universalizability and argues that it is not a case of universalizability at all. Carlson does so by examining the sports analogy, given that Kalin’s defense of egoism “depends for its plausibility on the analogy with the competitive game situation” (Carlson, 1973, p. 27). In highlighting a flaw in the sports model, Carlson exposes a similar weakness in Kalin’s claim that ethical egoism can be consistently universalised.

Carlson focuses his attention on the following passage from Kalin:

But does believing that A ought to *Y* commit one to wanting A to do *Y*? Surely not. This is made clear by the analogy with competitive games. Team A has no difficulty in believing that team B ought to make or try to make a field goal while not wanting team B to succeed, while hoping that team B fails, and, indeed, while trying to prevent team B’s success. Or consider this example: I may see how my chess opponent can put my king in check. That is how he ought to move. But believing that he ought to move his bishop does not commit me to wanting him to do that, nor to persuading him to do so. What I ought to do is sit there quietly, hoping he does not move as he ought (Kalin, 1995, p. 82-83).

According to Carlson (1973, p. 27), Kalin conflates the hypothetical '(I believe) A ought to *Y* if A wants to succeed' with the categorical '(I believe) A ought to try to succeed'. In the football example the fundamental belief, derived from universalising egoism, is not that A believes that B ought to try to make a field goal. Rather, what A believes is that B ought to try to win the game. In saying that 'team A' believes 'team B' ought to kick the field goal Kalin suppresses the antecedent; what team A believes is that team B ought to make or try to make a field goal *if* team B wants to win the game.

However, given that team A does not want team B to actually win the game (and winning the game involves successfully scoring field goals), team A "does not believe that team B ought to make a field goal" (Carlson, 1973, p. 27). Likewise in the chess game, it is not the case that (I believe) my opposition ought to put my king into check but that I don't want him to; rather (I believe) my opposition ought to put my king into check if my opposition wants to win the game. I do not want my opposition to 'succeed' in winning the game and therefore I do not believe my opposition ought to succeed at putting my king into check.

Kalin makes a two-part claim (i) I believe A ought to *Y*, and (ii) I don't want A to *Y*. However, once the hypothetical is exposed (ii) is rendered obsolete. Similarly, in the chess game, that I don't want my opposition to put my king into check is made redundant when the hypothetical 'A ought to put me in check if A wants to win' is exposed. While it is true that I don't want my opposition to put my king into check it is also true that I do not believe my opposition ought to succeed in putting my king into check given that I don't believe the opposition ought to win the game. The crux of Carlson's criticism is that Kalin has conflated (i) what '(I believe) others ought to try to do, but that I don't want them to do' with (ii) what '(I believe) others ought to do (i.e. to succeed at doing), but that I don't want them to do'. According to Carlson an egoist could hold (i) which equates to, others ought to attempt to be egoists but I want them to fail, but cannot hold (ii) others ought to be successful egoists. If (ii) cannot be held then universalisation (in Kalin's weak sense) has failed.

Presumably the reason I (Agent₁) act in accordance with the dictates of egoism is that I believe that the right thing to do is to make every attempt to maximise my interests. Therefore an egoist can be referred to as having successfully achieved the goals of egoism if at the end of the day the egoist's interests have, as a matter of fact, been maximised. As the egoist cannot believe an opposition 'ought to move so as to win the game *if* the egoist's interests are maximised by winning', likewise an egoist (Agent₁) cannot believe that 'others ought to successfully act in accordance with the dictates of egoism *if* in doing so the actions of others will prevent, or negatively impact upon, Agent₁'s interests.

If an agent holds:

(i) I (Agent₁) ought to be successful in following the dictates of egoism

then the correctly universalised form of (i) is:

(ii) I (Agent₁) want (desire) others (Agent_n) to successfully act in accordance with the dictates of egoism

and, if others acting as egoists will be detrimental to Agent₁'s self-interests, then Agent₁ also holds:

(iii) I (Agent₁) want (desire) "the circumscription of the range and multiplicity of [my] own satisfactions" (Carlson, 1973, p.28).

Such wants are clearly inconsistent and, ultimately, incoherent. The egoist wants both to succeed in maximising interests and, in wanting others to be successful, to fail at maximising those interests. As Carlson (1973, p. 29) notes, the only way Kalin can avoid this inconsistency is if he holds the position that an egoist believes others ought only to *attempt* to be egoists and that the egoist wants them to fail in this attempt. That is, to avoid inconsistency, Kalin's egoist must hold:

(i) the egoist (Agent₁) ought to successfully maximise self-interest
(ii) others (Agent_n) ought to attempt but fail at maximising interests.

However, the expense of retreating to such a position is the giving up of universal egoism in favour of individual egoism; (ii) is not a universalised form of (i).

§4.2.3 Private Universal Egoism

At this point it would appear Kalin's private universal ethical egoism is in serious trouble. Firstly, according to Carlson the attempt to universalise egoism has failed. Kalin conflates the distinction between attempting to act as an egoist and successfully acting as an egoist, and his universal egoism is really just an elaborate formulation of individual egoism. Secondly, even if Carlson's criticism could be overcome, I am extremely doubtful that a private morality is possible.

However, I will argue that Kalin's attempt to universalise egoism can be successfully developed and that the egoism that follows from such a development is both consistent and public – an agent can promulgate egoism, in most cases, without acting in a manner that is detrimental to the agent's self-interest. To do so it will be necessary to develop a two-factor account of egoism, one that separates reasons for action and motivation to act.

§4.3 Re-Universalising Ethical Egoism

The ultimate principle the egoist believes, is that if Φ is good in itself relative to the egoist's worldview then Φ provides the egoist with strong moral reason to act. However, that an agent who subscribes to egoism has *moral* reason to act for that agent's own self-interest does not lead to the conclusion that an agent who subscribes to egoism has an overriding reason to act in order to bring Φ about on each and every occasion; as discussed in §2.3, the egoist may have numerous reasons to act otherwise. In applying the principles of egoism in a specific scenario, the egoist will be advised to act in a specific manner, however the egoist may act nonmorally and instead apply some other theory of action guidance⁵³, a theory that may advise the agent to act in some other manner. If n valid theories of action exist and no specific theory can be shown to be objectively superior then the answer to the question which course of action will I take in this situation is multiply realisable. If egoist theory is applied then acting so as to bring Φ about is determined as the correct moral act. However, the egoist need not consider every decision he makes as a moral one and may apply a nonmoral theory in

⁵³While consistency requires that the agent ascribes to a single moral theory the egoist may also ascribe to other decision making theories (such as financial) that are more applicable within certain contexts.

determining how to act in a specific scenario.⁵⁴ It is only when the egoist considers the scenario has moral implications and asks ‘how ought I to act?’, that acting in a manner determined by a nonmoral theory and contrary to that advised by egoism is determined as less than ideal. An egoist does not claim that certain types of reason are objectively superior. However, if the egoist determines this to be a moral scenario, such that the question ‘how ought I to act’ is relevant, then in doing so the egoist is giving preference to reason for acting based on the egoist axioms of self-interest.

Universalising this principle does not require incorporating the claim that egoism is overridingly superior to other theories and therefore universalising egoism does not determine how an agent who subscribes to egoism *must* act (as Medlin, Kalin, and Carlson seem to accept) but rather provides moral reasons for acting as well as the moral justification for those actions.

If Φ is recognised as good in itself relative to Agent_n’s worldview then Φ provides Agent_n with strong moral reason to, and justification for acting to bring Φ about.

This first factor principle provides an egoist with moral reason to act but does not in itself completely motivate the egoist to act, as competing reasons may motivate the egoist to act otherwise. In making such a claim I am accepting that reasons are causes of actions in something akin to Davidson’s position:

R is a primary reason why an agent performed the action A, under description d, only if R consists of a pro-attitude of the agent towards actions with a certain property, and a belief of the agent that A, under the description d, has that property. (1963, p. 687)

An egoist has a pro-attitude towards self-interest and will find reason for acting where the agent believes that a certain action will hold that property. However, an agent may hold several such pro-attitudes that may provide the agent with competing reasons to act in different manners. That said, the egoist places strong

⁵⁴ As indeed we often do. We do not, as a rule, morally question how we ought to act with regard to day to day routine. How much will I spend on groceries is usually the result of financial decision making rather than based on moral concerns.

import on the moral doctrine of egoism and the pro-attitude toward self-interest is likely, all things considered, to provide the egoist with primary reason for acting.

I began this chapter assuming that if an egoist ($Agent_1$) recognised Φ as $Good_{w(Agent_1)}$ then $Agent_1$ had strong reason for bringing Φ about and that it was also reasonable to assume that $Agent_1$ desired Φ came about. As has already been noted the egoist may have other reasons to act, reasons that are the result of applying nonmoral theories of action and that may prescribe a different course of action. Further, given the complexity of human life it is likely that several outcomes may also be recognised as $Good_{w(Agent_1)}$ and that both Φ and Ψ may be $Good_{w(Agent_1)}$ and provide $Agent_1$ with strong mutually exclusive reasons to act. These problems can be overcome, to some degree, by adding the clause ‘and would maximise an agent’s interests’ to the egoist principle:

If Φ is recognised as good in itself relative to $Agent_n$ ’s worldview and acting to bring Φ about would maximise $Agent_n$ ’s interests then Φ provides $Agent_n$ with strong moral reason to, and justification for acting.

This overcomes the problem of competing goods. However, that an agent has strong moral reasons to bring Φ about does not necessarily lead to an agent wanting or being motivated to bring Φ about as, on a specific occasion, an agent may be motivated to act from reasons that arise through the application of nonmoral theories. There is nothing irrational in claiming that ‘egoism is the preferable moral theory but on this specific occasion I am going to preference nonmoral reasons for action’, for example an agent may simply act on a whim.

The egoist’s ultimate moral principle is that maximised self-interest provides an agent with reasons for acting. This principle can be universalised without contradiction or inconsistency and simply points to the uncontroversial concept that all agents have self-interested reasons for acting⁵⁵. It is a completely different matter whether an agent will, or should, act because of those reasons.

⁵⁵ It is of course controversial that these are moral reasons. But this is secondary to the point I am currently making.

§4.4 Ethical Egoism and Moral Motivation

As I argued in Chapter 2 it is unlikely moral reasons have overriding status, therefore the degree of motivation an agent will have to act morally depends on the degree of conviction an agent has that reasons derived from the application of a moral theory are preferable to reasons derived from a nonmoral theory or principle. I contend that an *ethical* egoist is someone who:

- (i) believes that in most cases moral theories provide reasons for action that are preferable to nonmoral ones.
and
- (ii) believes that moral reasons derived from self-interest are preferable to those derived from other-regarding reasons for action.

I have already argued that so long as a theory meets some minimum requirements (consistency, coherence, possibly noncontradiction) then it is not logically possible to claim that any specific theory that meets these minimum requirements is superior to any other theory that also meets these minimum requirements (§2.3)⁵⁶. Ultimately the method of moral reasoning that an agent subscribes to is entirely individual and a matter of personal preference (assuming, as I have claimed, that more than one method of moral reasoning meets these requirements). That an agent believes a theory X, given that X is a coherent consistent theory, offers the best explanation for Y is, in itself, suitable justification for an agent subscribing to X. To convince someone that they are mistaken in this approach requires either convincing them that another theory is preferable (for whatever reason) or demonstrating that the currently preferred theory fails to meet the minimum requirements for being considered a theory of any kind.

For an agent to claim that a theory, in this case egoism, is preferable (in that it provides better reasons for action than do other theories) is also to accept that, in most cases, those reasons provide the agent with the motivation to act in the way that the theory ascribes. In short, if A believes X is the preferable theory of action guidance and X recommends acting to bring Φ about then A is likely to be motivated to act so as to bring Φ about.

⁵⁶ An extra criterion is necessary if we are to separate moral theory from theories in general. That a theory is coherent etc is insufficient for claiming it provides a relevant solution to a moral question. I will examine this final criterion in the final chapters of part one.

§4.4.1 Internal and External Reasons

The position I am outlining has some similarity with Bernard Williams' (1981) thesis that "we cannot have genuine reasons to act that have no connection whatever with anything that we care about" (Chappell, 2006). However, I am not claiming that external reasons for action (reasons not personally connected to an agent's wants and desires) can always be reduced to some kind of internal desire satisfaction, except perhaps on some purely formal level. At a very basic level it appears to be part of the human condition to try to make sense of the world and thus, at a primary level all reasoning is fueled by an internal desire for explanation.

However, making sense of the world requires not only explaining events in the natural world but also explaining why humans act as they do. In this respect when we question why X did Y we are asking for greater explanation than just X had an internal desire to Y and doing Y satisfied that desire. What we are often asking is, what reasons provided the dominant motivation for X doing Y?

Every agent, prior to acting, may face a multitude of competing reasons. When contemplating 'What ought I to do?' an agent finds external reasons of the form: Financially I ought to..., legally I ought to..., for other-regarding reasons I ought to..., for self-interested reasons I ought to...⁵⁷, all of these reasons, in as much as they are the result of consistent and coherent theories of action, equally present the agent with reasons to act. However, and in this I agree with Williams (1981, p.111), none of these reasons in themselves motivates the agent to act. While initially it is an agent's desire that motivates an agent's rational deliberation over 'how the agent ought to act' (without a desire for answers the deliberation process would not even commence), it is from that point a continual synthesis of internal and external reasons that lead to, and provide, an agent's desires and motivation for action.

Human desire for answers leads to the deliberation over external reasons for action. Deliberation over external reasons for action leads to an agent formulating

⁵⁷ Ethical self-interested reasons for action are external and the result of rational considerations whereas physiological egoism would represent internal reason for action.

a belief that X is a better type of reason for action⁵⁸. A belief that X is a better type of reason provides internal reason (and motivation) for acting upon some specific external reason. Beliefs about the best type of reason lead to deliberation over, and preferencing of, certain types of external reasons for action in specific contexts. Such a process is not static; in deliberating over external reasons in specific contexts and debating these reasons, agents may, over time, come to modify their beliefs and preferences over what represents the best type of reason (or, if a type of reasoning can be shown to be unsound, to dismiss that type of reasoning altogether).

To recapitulate, in attempting to answer the question ‘how ought I to act?’ an agent may formulate the belief that actions that accord with reasons of the type provided by the egoist doctrine are preferable. Motivated by the belief that egoist reasons are the preferable type of reason, an agent who faces a specific moral dilemma is likely to be motivated to act in accordance with the egoist doctrine. However, if the act promoted by the egoist doctrine conflicts with the agent’s other beliefs (other reasons for acting or, perhaps, intuitions) the agent may find it necessary to re-examine the initially accepted theory. If that agent then finds it necessary to modify or replace the previously accepted theory then a new belief about the preferable type of reason for action will be formulated and along with it the motivation to act in such a manner.

Reasons for acting are, and must be, universalisable. However, that a reason to act provides an agent with the motivation to act on those reasons depends upon the agent giving preference to those reasons over competing reasons that might also motivate action. The ultimate principle of egoism provides reason for acting and is universalisable. Accepting the dictates of egoism as providing reasons to act is a matter of assigning individual preference to those reasons.

- (i) (*Universal principle of ethical egoism*) If Φ is recognised as good in itself relative to Agent_n’s worldview and acting to bring about Φ is the act that would maximise Agent_n’s interests then Φ provides Agent_n with strong moral reason to, and justification for acting.

⁵⁸ It might also lead to the agent adopting some nihilistic position that no reasons are preferable. However, I take the very fact a nihilistic position frustrates the initial desire for answers to be one of the reasons the position is avoided.

- (ii) (*Individual motivation for acting*) If Agent_n believes that ethical egoism provides preferable reasons for acting then Agent_n will, in most cases, be motivated to act in accordance with (i).

The egoist principle (i) is universal in that every agent is provided with self-interested reasons to act. However, the motivation for acting (ii) is individual as the belief and acceptance of a specific theory as preferable is relative to the individual agent. Egoism might appear paradoxical if it is seen as dictating how an agent must act. But this is not actually the case for once it is accepted that egoism is simply a theory that provides reasons for acting it can be seen that although an agent believes that X is the preferable theory this does not equate to X being the preferable theory, nor that all agents ought to act in accordance with X.

To recapitulate, I am examining the claim that egoism is a coherent, consistent and non-contradictory theory of action. I am not claiming it is the theory of action or that it is inherently superior to other theories of action. As I have previously argued it is logically impossible to label one of competing coherent theories as superior except from within the viewpoint of that theory.

§4.5 Ethical Egoism and the Sports Model

Kalin's defence of egoism (§4.2) relies on an analogy with a sports model, a model with which I am uneasy. It simply does not seem to represent a good analogy given that games, in general, are built on a model of conflict whereas morality is supposedly concerned with conflict avoidance and resolution. A good competition is one where the state of conflict is present and promoted throughout the duration of the game. Morality (generally) is designed to resolve conflict and in terms of morality the sooner the conflict is resolved the better. It would appear that applying moral theory to a game scenario would effectively be to defeat the purpose of the game. Consider the bizarre results of applying utilitarian logic to the sports model. Should I kick a field goal? Well, if my team are the unsupported underdogs and more people would gain satisfaction from the favourites winning then No! I ought to refrain from kicking, or purposely miss the field goal.

However, for consistency with the previous sections I will retain the model and

while I am doubtful of the analogy it does at least do enough to demonstrate that egoism can be universalised without any of the inconsistency that Carlson envisages. Consider the final moves in a game of chess.⁵⁹ I have just made a bad move such that if my opposition (Deb) notices my error she will be able to capitalise on the mistake, move her rook, place me in checkmate, and win the game. However, if Deb overlooks my mistake it is I who will be able to capitalise on the position and win the game. If the egoist principle is applied to this case it would seem that:

- (i) Upon noticing my error, and assuming Deb wants to win the game, Deb has strong reason to move the rook and win the game.
- (ii) I, in wanting to win the game, have strong reason for wanting Deb not to move the rook.
- (iii) I, in wanting to win the game, have strong reason for not drawing Deb's attention to the game-winning move.

There is nothing inconsistent in my accepting that Deb has good reason to place me in check (universalised ultimate principle of egoism) and my also not wanting Deb to place me in check (I lack motivation toward this end), hoping that either (a) Deb doesn't realise such an act is an option, or, (b) is motivated to act from some other reason. Universalising reason for action does not lead to wanting those actions to come about or to wanting others to want those actions to come about. Neither am I prevented from promulgating the egoistic principle and discussing reasons for action; Deb would undoubtedly be completely unsurprised, after asking what she ought to do, by my stating the somewhat obvious claim that she has good reason to make whichever move is most likely to promote her winning the game. In terms of motivation to act, (in so much as I am an egoist) I am motivated by my belief that egoism provides a better type of reason to accept (ii) and (iii) as preferable reasons in this scenario and I am motivated (iii) not to draw Deb's attention to the rook move.

However, a problem arises if Deb asks 'Should I move the rook?'⁶⁰ How, as an egoist, am I supposed to answer this question? I would be answering truthfully if I replied 'No, although you have good reason to move the rook, in that it would

⁵⁹ I selected the chess game analogy as this simplifies the event to competition between two agents rather than two teams.

⁶⁰ A possible but unlikely state of affairs in that if Deb wants to win, and recognises the rook move, then this in itself ought to motivate Deb to move the rook.

further your self-interest in enabling you to win the game, I don't want you to be motivated to act for this reason as it would result in the thwarting of my self-interest'. However, in giving such an answer I would expose my vulnerable position and, in all likelihood, give the game away contrary to the dictates of egoism.

I will shortly return to how I think the egoist ought to analyse such a scenario. Firstly, I want to consider the type of answer Carlson would likely contend the egoist ought to give. Carlson would envisage a problem for the egoist in answering Deb's question. The egoist, in holding the belief that the egoist principle provides the preferable means by which to answer questions regarding how an agent ought to act, should also hold the "disposition to induce in others attitudes consistent with what one believes" (Carlson, 1973, p. 29). If Carlson's claim is correct then it seems the egoist answer to Deb's question ought to be 'Yes, you have good reasons to move the rook and the reasons for doing so are preferable to any reason you might have for acting otherwise', but to give such an answer is to act contrary to the egoist's best interests. It seems that the two-factor version of egoism has simply pushed the inconsistency problem back a step. The egoist can both universalise the egoist principle (reasons for action) and promulgate this principle (each agent has egoist reasons for action). However, the egoist cannot promulgate that egoism is the preferable reason for action without being inconsistent and acting in a manner detrimental to the egoist's interests.

Carlson (1973, p. 29) does not assume that an egoist must act so as to induce consistent attitudes in others on every occasion but "only when the appropriate circumstances obtain". Sports and games, which are built on models of conflict, are simply not one of those occasions. The egoist can promulgate the egoist principle on most occasions given that most are well aware that self-interested reasons for action exist. However, the interesting question is whether 'appropriate circumstances' exist such that, in those circumstances, an egoist ought to act so as to induce a similar attitude in others and where, in doing so, the egoist would be acting in a manner detrimental to the egoist's interests.

In most everyday settings I can see no reason why an egoist must attempt to convince others that the egoist's preferred theory is generally preferable except, possibly, in circumstances where the egoist is required to justify an action taken. One scenario where an egoist would be required to promulgate and expound upon the preference toward egoism would be in serious philosophical moral debate. However, in this context there appears to be no reason why the egoist cannot promulgate (i) the universal principle of egoism and, (ii) the reasons why self-interest is a preferable principle upon which to build a moral theory. Moral debate is theoretical. No actions are being chosen as a direct consequence of the debate and the discussion itself is unlikely to have any detrimental effects on the agent expounding the theory.

As Medlin claims, ultimate principles must be subjected to rigorous testing, must be universalisable and public. However, while we can insist that theories, if they are to be classified as theories, meet a minimal set of criteria such that they are coherent and consistent, we can do little more as it is logically impossible to determine an objective criterion by which to judge between those theories that meet these minimum requirements. At the end of the day, the selection of one theory over another *is* a matter of preference. There is no fact of the matter as to the correctness of preferences and, in acknowledging this, no reason why I must promulgate my beliefs and attempt to sway the preferences of others.

Returning again to the chess game, I take it that all of the following reasons for acting and motivations to act come into play:

- (i) [By universalizing the egoist principle] The egoist *must* accept that Deb has good reason to move the rook.
- (ii) [By egoism] The egoist has good reason for wanting to win the game.
- (iii) [By egoism and (ii)] The egoist has good reason for wanting Deb not to move the rook.
- (iv) The egoist believes that the egoist principle provides the preferable means of moral reasoning.
- (v) The egoist is motivated by (iv) an internal preference for egoism, and, (ii) the external reason of wanting to win the game and can either (a) not answer Deb, (b) offer a reply in terms of (i) that Deb ought to move the rook if she believes it to be the best move, or (c) lie.

The ultimate principle of egoism is that an egoist has reason to maximise interests. I imagine that (c) is unlikely to maximise interests as the lie (which is likely to be uncovered) will probably have further repercussions that outweigh any gains made by winning a single game of chess. Sitting mute (a) is, for obvious reasons, unlikely to have productive results. It would seem that maximising an egoist's interests in this chess game requires that Deb's question be answered by 'if you think it is the best move'. The belief that the doctrine egoism provides a better type of reason together with the ultimate principle of egoism (reasons to maximise interests) motivates an egoist, on this occasion, not to promulgate the advice to move the rook.

§4.6 Conclusion

Medlin's criticism of egoism was that attempting to universalise egoism led to general inconsistency to the point where the theory became incoherent and offered no moral guidance. I have shown Medlin's criticism to be unsuccessful in that it conflates the egoist ultimate principle (reasons to act) and the egoist's motivation (reasons for acting). The ultimate principle of egoism provides reasons for acting but it does not obligate the egoist to act. There is no inconsistency in universalising the principle of egoism, which is simply that:

If Φ is recognised as good in itself relative to Agent_n 's worldview and acting to bring Φ about is the act that would maximise Agent_n 's interests then Φ provides Agent_n with strong moral reason and justification for acting.

Whether an agent acts on these reasons depends on the individual beliefs and preferences of the agent and these beliefs and preferences cannot be universalised.

If Agent_n believes that ethical egoism provides preferable reasons for acting then Agent_n will, in most cases, be motivated to act in accordance with the universalised form of egoism.

The perceived inconsistency in egoism occurs only when the egoist's reasons for action are interpreted as determining how the egoist *must* act. However, while the criticisms examined in this chapter have been defeated this is not enough to save egoism from the charge of being inconsistent. If, in a specific context, an agent decides to apply the principles of egoism (believing egoism to be a theory of action guidance) then egoism, if it is a coherent theory, ought to provide the agent

with action guidance that has a reasonable probability of realizing the agent's goal of maximizing self-interest. I will now turn my attention to the question of whether egoism can provide such action guidance.

Chapter 5

Further Inconsistencies & Incoherence

§5.0 Introduction

Moral theory as a subject investigates the nature of good. According to egoism the nature of good is the maximisation of self-interest. Moral theories also provide action guidance, that is they attempt to answer the question ‘How ought I to act in this specific situation?’ According to egoism the answer to this question is that an agent ought to act so as to maximise self-interest. The criticisms discussed so far have, for the most part, assumed that good and self-interest are mutually exclusive, an assumption that I have shown to be false.

In this Chapter I will consider two types of criticism that aim to show that a theory of action guidance built on a principle that equates good with self-interest is inconsistent, incoherent, or both. While each of these criticisms might have been extended to fill an entire chapter, space considerations required that I impose some limitations on the discussion.

The first criticism that will be examined in this chapter is the claim that, following the prescriptions of egoism (acting so as to maximise self-interest) prevents an agent, in some contexts, from attaining the maximisation of self-interest. This criticism is similar in form to the hedonistic paradox, (where the continual pursuit of happiness prevents the attainment of happiness), but somewhat more complex in nature. It is claimed, in prisoners dilemma type contexts, that game theory shows it to be logically impossible for an agent to attain the goal of maximising self-interest by following the dictates of consequential egoism.

I will begin this chapter by examining the prisoners’ dilemma in its simplest form, a scenario that presents the egoist with a single choice (a one-off ‘game’), and will then demonstrate that it is far from obvious that an egoist would adopt the strategy of *always defect*⁶¹. Gauthier (1974) uses game theory to analyse the actions promoted by egoism and claims that an agent ascribing to egoism is required (in some contexts) to act in a manner that prevents the agent from acquiring the goal

⁶¹ Always defect is one of many game theory stratagems, as will become clear as the chapter progresses.

prescribed by egoism. It is difficult to engage with such criticisms without moving deeply into the domain of game theory. However, I will engage sufficiently to demonstrate that Gauthier's game theory analysis of the prisoners' dilemma is incorrect; the dilemma cannot be realistically treated as a single event non-zero game. When the context of the dilemma is examined, and necessary iterations are considered, the paradox of the dilemma disappears and egoism appears completely rational. Secondly, I will contend that game theory does not easily apply to the real world and that the outcome of artificially constrained and restricted games tells us very little about how an agent will, or ought to, act in what might appear to be an analogous real life situation.

The second type of incoherence criticism I will look at regards the nature of the self. The self is a central concept within egoism as the target (beneficiary) of egoist actions is the self. If the self cannot be demarcated then the egoist faces a serious problem; the target (beneficiary) of an egoist's actions cannot be determined. Further, the criterion for determining the moral status of an egoist's actions dissolves as it cannot be determined whether an action maximises *self-interest*, unless there is a specific self to whom those interests can be said to belong. Zemach (1978) advances this line of reasoning, claiming the notion of the self is nothing more than convention and that the claim of acting out of self-interest is incoherent. This line of criticism is not limited to the physical extension of the self. If the egoist's claims involve a balance between current and future interests then it also needs to be shown that the notion of a self existing over time is coherent. Parfit (1984), amongst others, has claimed the self is actually nothing more than a series of psychologically related life stages, and the notion of a self is vague and refers only to a limited temporal self (that is, a life-stage that is part of but does not stretch across a full biological lifespan). Parfit claims that the relationship between these temporal-selves is not transitive (such that the connection between distant life stages may be lost). If such claims are true then an egoist who claims to be acting out of self-interest when the target of those actions exists in the relatively distant future is not acting so as to benefit him or her self. Rather, the egoist is acting as a limited and misguided altruist, and for the benefit of some other.

While Parfit brings to our attention alternative notions of the self that do not necessarily exist from birth until death, unless it can be shown that the concept of a self that stretches across a biological lifetime is incoherent then an egoist is not compelled to abandon such a claim. Given legitimate alternatives for explanations of the nature of the self the egoist has good reason for accepting as the best explanation a self that extends across a biological lifespan. However, in doing so I will expose egoism to the final challenge of this section, namely, that the theory of egoism is an unjustifiable hybrid that merges agent relativism with temporal neutralism. This criticism is based on the assumption that all justifiable theories must be pure, that is, they must not mix methodologies, and that moral theories must be either entirely relative or entirely neutral and cannot be arbitrarily mixed.

§5.1 The Prisoners' Dilemma

The prisoners' dilemma is a situation where the reward or punishment you will receive is relative to and affected by the actions of another. The specific dilemma, as introduced in §1.2.1 was:

Two known criminals have been arrested on suspicion of committing a crime. The evidence against both of them is quite weak and unless the guards can elicit a confession they will both receive only relatively small jail sentences. The guards approach each suspect separately and offer the following deal:

If you confess (and in doing so provide evidence against the other), and the other prisoner remains silent, then you will be set free and the other will face eight years in jail (and vice versa).

If you do not confess, and the other prisoner remains silent, then the evidence is sufficient that you will both face two years in jail.

If you both confess then you will both be jailed for five years.

The two agents (who for convenience I'll refer to as Alfred and Brian) each has two and only two options, confess or stay silent. Depending on the actions of the agents one of four possible outcomes could occur. Alfred confesses and Brian does not, Brian confesses and Alfred does not, both Alfred and Brian confess, or neither Brian nor Alfred confesses. The first two of the four are different actual outcomes but are both of the same type, one confesses and one does not. If game

theory is applied to this dilemma then the utility of the possible outcomes is as follows:

		Brian	
		Confess	Don't Confess
Alfred	Confess	(iii) -5,-5	(i) 0,-8
	Don't Confess	(iv) -8,0	(ii) -2,-2

(i & iv) Either Alfred or Brian confesses and the other does not. As a result the confessor goes free (a utility of 0) and the other faces 8 years in jail (a utility of -8). I am assuming a year in jail equates to a utility of -1 and that these are the only factors relevant to the calculation of utility for this act.

(ii) Both Alfred and Brian stay quiet and both serve two years in jail. The utility, to each individual agent is -2.

(iii) Both Alfred and Brian confess to committing the crime and as a result they both serve five years in jail, a utility to each agent of -5.

It is usually contended, taking the viewpoint of Alfred, that if Alfred is an egoist then his reasoning in this dilemma will be along the following lines:

1. If Brian confesses and I stay quiet, I will face eight years in jail (-8).
2. If Brian confesses and I confess, I will face five years in jail (-5).
3. (by 1 and 2) If Brian confesses my best course of action is to confess.
4. If we both stay quiet, I will face two years in jail (-2).
5. If Brian stays quiet and I confess, I will not be sentenced (0)
6. (by 4 and 5) If Brian stays quiet my best course of action is to confess.
7. (by 3 and 6) Whatever Brian does I am better off if I confess.

Brian will reason the same way and in attempting to maximise their individual self-interest, Alfred and Brian will both confess. However, in doing so they fail to maximise their self-interest and both serve five years in jail whereas by remaining quiet they would have served only two years in jail.

This leads to the criticism that egoism is paradoxical (and thus ultimately incoherent) in that (in some contexts) following the dictates of egoism will prevent an agent from maximising self-interest and hence a theory of right action

will prevent an agent from having acted rightly. If Alfred and Brian determine how they ought to act by applying the principles of egoism (act so as to maximise self-interest), the outcome will be that they both fail to maximise self-interest. It would seem that, within this context, maximising self-interest requires that the agents do not act from self-interest.

I should elaborate a little at this point. The criticism is not just that there exists *a* scenario where applying a theory of action guidance prevents the outcome that the theory is designed to obtain, but that game theory offers mathematical proof that, if an egoist theory of right action is adopted, then in any scenario that pits egoist against egoist, in such a manner that the egoist's action is based upon minimising the effect of the other agent's actions, then the maximisation of self-interest becomes impossible. If we do face prisoners' dilemma type situations in the real world then the adoption of egoism will lead to non-optimal social instability; a state of social conflict where maximisation of self-interest is reduced to avoiding risk at the expense of everyone being worse off.

Assuming that situations analogous to the prisoners' dilemma do exist, it is still far from clear that the game theory analysis of how the egoist is required to act is correct. The prisoners' dilemma has been treated as if it were a simple single payoff calculation, which requires that the game begins in the cells and ends once the decisions have been made, that is, that no other interactions are relevant to the decision about to be made. While the prisoners' dilemma, as described, limits the agents' choices to one of two alternatives I contend that it is highly improbable that such dilemmas, in the real world, can be further restricted to a single occurrence event. For such dilemmas to be treated as single occurrences would require that the two agents had no previous interactions and, in all probability, would not be involved in any future interactions. Setting up the prisoners' dilemma as a single event, as portrayed by those criticising egoism, would require that:

- (i) Alfred and Brian had only just met and, with extremely little knowledge of each other, had decided to commit a crime. This ensures the game begins in the cells.

- (ii) It is highly unlikely than Alfred and Brian would ever meet again after the decision of whether or not to confess had been made. This ensures that the game ends when the decision has been made.

It is of course possible to create such a scenario that meets these criteria; each of the criminals might have received his instructions from a third party and never met the other. However, the more clauses that are introduced to ensure that the prisoners' dilemma is best described as a single one-off event the less likely the event could actually occur in the real world. The clause just suggested requires that an egoist would decide to embark on a risky criminal venture, blindly trusting an unknown accomplice in a manner that seems directly *against* the prescriptions of egoism; the risk involved is simply too great and those advancing the prisoners' dilemma scenario implicitly accept that the egoist will always act so as to avoid such risks. If, on the other hand, the egoist had reason to trust the other agent (which would suggest previous interaction) then those who advance the criticism would have us believe that the criminals trusted each other enough to enter into a risky criminal venture but, once in the cells, do not trust each other enough to remain quiet. That is, they would trust each other at least to the degree that they would act professionally, would not betray the other, and would share the illicit gains, but that as soon as they were in the cells all such trust would vanish and each egoist would accept no risk, and instantly turn on his accomplice.

The scenario could be further altered, but the result of continually modifying the scenario (such that it qualifies as a single event and is such that egoists could find themselves in such a position) results in the dilemma dissolving. For example, perhaps the egoist had come to the conclusion that the risk involved in trusting another was commensurate with the rewards to be gained from successful completion of the criminal endeavour. The two criminals may then find themselves in the scenario and facing the prisoners' dilemma. Further, perhaps the reward for confession included a new fake identity (such that no new interactions between the two were likely). If all these conditions are met then the scenario is as the critics describe and the two criminals face a one-off decision. However, if this is the case the decision to confess is anything but irrational. Alfred in confessing, acts so as to have the best probability of maximising his interests. If both confess then the failure to actually maximise interest is not due to a fault with egoism, or

due to irrationality, but to having insufficient information. If there is a problem here it is not with egoism but a more general problem for all consequentialist theories, namely that rightness depends to a large part on factors beyond the agent's control and that are often unforeseeable.

However, I remain dubious situations that can be classified as one-off events and place egoists at loggerheads are likely to occur in the real world. It seems more likely that events preceding the situation and the (probable) events that will follow are of relevance in solving the dilemma. I will contend that, when making the decision whether to confess, the rationale for Alfred's action, even if Alfred is fully aware that Brian will also act so as to maximise self-interest, will be different than is usually assumed and that Alfred will not necessarily confess.

If it is accepted that the prisoners' dilemma cannot be treated as a standalone event and that the two criminals have had some past interaction (at the very least in deciding to commit a crime) and further, that they are likely to have future interactions, then a single- event matrix cannot be used in determining how Alfred and Brian ought to act. While the game theory strategy of *Always Defect* (in this case confess) is, from a single agent's perspective always the superior strategy within the context of single events, this strategy fails when iterations are considered. I contend that when iterations are considered the action promoted by the egoist doctrine is to stay quiet and that both criminals, even realising that the other is likely to stay quiet, will find it to be in their self-interest to do likewise.

§5.1.1 Iterated Game Theory

As Poundstone (1993, p. 237) correctly notes, "when only the present dilemma is subjectively important, it is effectively a one-shot dilemma in which defection is to be expected. But when players value future as well as present gains, a true iterated dilemma exists." I have already argued that an ethical egoist is one who considers balanced short- and long-term gains and is not a hedonistic presentist. Further, whichever action Alfred and Brian take it is highly probable they will have future interactions, either in prison (should they both stay quiet or both confess) or some years later, if one confesses and one does not, when a (likely vengeful) cohort is finally paroled.

The best strategy of action in an iterated prisoners' dilemma is a contentious issue, one that is still far from resolved and is subject to ongoing research within the field of game theory. In 1980 Robert Axelrod created a series of computer tournaments to examine strategies for solving the prisoners' dilemma. This tournament took the following format; game theorists submitted 'systems of action' (the strategy of *always defect* being one such system⁶²), each system was then pitted against every other system for a sequence of two hundred iterated dilemmas. In each iteration points were awarded as follows: mutual cooperation gained 3 points, defecting where the other did not gained 5 points, mutual defection gained 1 point, and cooperation where the other defected gained none. The values are a little different from those assigned to the Alfred/Brian prisoners' dilemma but the resulting matrix is effectively the same.

	Cooperate	Defect
Cooperate	3,3	0,5
Defect	5,0	1,1

The highest possible score from this tournament would be 1000 points where the winner defected on every occasion and the opposition cooperated on every occasion; a score that was gained in a single playoff when the *always defect* strategy faced the *always cooperate* strategy, but a score that was impossible across the entire tournament. *always defect* scored only 200 points when it played against an identical strategy. A more realistic score ceiling would be around 600 points and this could be attained by averaging three points every round (the score assigned to continual mutual cooperation, but is unattainable by a strategy of *always cooperate*, which suffers when facing *always defect*).

The strategy that prevailed at the tournament is known as *tit for tat* (which scored an average of 504.5 points). The *tit for tat* strategy cooperates unless the

⁶² The systems of action ranged from simple instructions such as *always defect* or *always cooperate* to complex patterns of action that involved lengthy lists of instructions

opposition defected on the previous round in which case *tit for tat* defects until the opposition once again begins cooperating. *Tit for tat* is, at first glance, an odd system because it “can’t beat anybody [i.e. any other system], but it still wins the tournament” (Axelrod cited in Poundstone. 1993). *Tit for tat* cooperates when others cooperate, gaining no advantage but scoring the same as if the strategy had been mutual cooperation. Similarly, *tit for tat* defects when up against *Always Defect* thus reducing the initial gain made by *Always Defect* and limiting its own losses to the 5 points lost in the first round.

I am not suggesting that an egoist must embrace a *tit for tat* strategy to life only that, if egoism is to be criticised by way of applying game theory analysis to a prisoners’ dilemma, it first ought to be determined which game theory strategy most closely matches the egoist’s ideals. Given that self-interest is the underlying principle of egoism and *tit for tat* is the strategy that wins (gains the highest score over time, and thereby maximises utility for that player over the length of a tournament) then it seems reasonable to assert that the egoist would subscribe to a *tit for tat* strategy.

§5.1.2 Prisoners’ Dilemma, the Egoist, and Tit For Tat

If, as seems likely, Alfred and Brian have a past, then unless they are acting out of revenge, their self-interest lies in cooperation. They each know the other would do better in the short term by defecting (so long as the other didn’t). However, they are also fully aware of the situation they will be in if they both defect which is more time in jail. Further, they will likely face some form of revenge if one defects while the other stays quiet. Mutual trust, based on past history, appears to offer a better chance of maximising interests than a possible, one-off gain, from defecting.

I suggest that Alfred’s egoist reasoning will be along the following lines:

1. While it would be best for me if I served no time in jail, if I confess and Brian does not then in eight years (perhaps less with parole) I will likely face an ex-convict looking for revenge.
2. Given that revenge is an attempt to balance the books or to gain an advantage, there is a reasonable probability that in the long term no advantage will have been gained by confessing.

3. Further, if I confess and Brian does likewise then we will both be sent to jail for an extended period. However, as a result I will also know that I cannot trust Brian and vice versa and this will likely be to our mutual detriment. Being aware that the other is playing a strategy of *always defect* will result in us both continually adopting this strategy to minimise losses.
4. [By 1,2,3] I ought not to confess.
5. *Tit for tat* appears to be the best strategy for maximising outcomes.
6. Brian has not previously defected (such as setting me up during the crime) and so the strategy of *tit for tat* is not to defect.
7. Brian ought to follow this same reasoning.
8. [By 5,6,7] I ought not to confess.

Alfred ought not to confess and while in not confessing he opens himself up to considerable risk he does so only because the alternative is worse and because the adopted strategy has the highest probability of maximising his self-interest. *Tit for tat* stratagems have their share of problems, the worst being that once one person transgresses against you an endless and unbreakable cycle of revenge commences. The second problem is that it requires two or more agents to view their past history in the same manner, a problematic state of affairs, given the numerous disputes from history where both sides adamantly claim that revenge was justified based on the action of the other. As Poundstone (1993, p. 253) notes, “The practical difficulty is not so much in knowing when to cooperate or to defect but to decide what is going on. In the real world it is not always obvious whether someone has acted cooperatively or defected. Actions can fall somewhere between the two extremes.” While I am not claiming that *tit for tat* is the best strategy, I am claiming that a strategy of *always defect* definitely isn’t. Real life situations are, for the most part, too complex to be solved by the application of simple stratagems. However, if game theory is to be used to analyse a situation then the *tit for tat* strategy appears to most closely match the egoist decision making process.

While it is true that Alfred or Brian would maximise self-interest if they could confess while the other stayed quiet, and that the gain made by confessing outweighed any future risk of revenge, then the egoist ought to confess. Within such a scenario, confession would maximise the self-interest of the confessor. However, it is hard to imagine such a scenario occurring and even if it did, given the guarantees that would need to be in place (to reduce risk to the degree that it ensures maximisation of the agent’s interest), it would not be a prisoners’

dilemma. It appears that in most cases the greatest chance for an egoist to maximise self-interest within the confines of game theory is to adopt a strategy of *tit for tat*. I contend that, in everyday life, ascribing to the theory of egoism does not prevent an egoist from attaining the goals of interest maximisation. The paradox of egoism arises only within one-off game situations, which outside of thought experiments and computer simulations are most unrealistic.

§5.1.3 A Final Word on Game Theory

One of the problems with game theory is that, “the nature of the game pre-empts other possibilities. The sentences are fixed, the choices are fixed; whilst this applies to the two prisoners, it is not obvious that every-day life generates such limited and limiting choices” (Mosely, 2001, para. 10). While a thought experiment is possible where two criminals are offered the outlined deal and, by a matter of stipulation no other alternatives exist, it is hard to imagine such a scenario actually occurring. Iterated prisoners’ dilemma situations do occur but they are not usually restricted in such an artificial manner and if those involved can enter into discourse then the strategies for dealing with the situations become increasingly more complex. At the very least a meta-game may become apparent. As Howard (1966, cited in Poundstone) claims, it is not just a case of whether someone defects but the reason why they defected. We might further question whether the defection is ruthless and deserving of revenge, or nice (in as much as the person defected only through a lack of options or some other mitigating circumstance) in which case we might overlook a previous defection and not extract revenge, thereby allowing an agent to break the endless cycle of revenge acts that are usually associated with the *tit for tat* strategy.

It may be that I’ve been too hasty and that it can be shown that agents do face single one-off type decisions as formulated in the prisoners’ dilemma, so I’ll return very briefly to considering the dilemma as a single action event. If this is the case then, rationally, the act that maximises an agent’s interest is to defect and to sacrifice what is often referred to as ‘the common good’. However, as Shubick (1982) noted, we should be no more surprised by this than we should be by learning that a feather and lead ball fall at the same velocity in a vacuum. Many philosophers’ moral intuitions may suggest that cooperation is the correct course

of action but in this case the explanation “is simply that common sense is wrong” (Poundstone, 1993, p. 277), just as the intuition that a lead ball will fall faster than a feather is wrong. The paradox of the prisoners’ dilemma does not exist; rational players in a one-shot, no compromise dilemma simply ought to defect because mutual cooperation in such a context is not a viable option. However, I remain unconvinced that actions can be isolated in such a manner and contend that the analysis of egoist actions as one-off events is seriously flawed. Iterated game theory is the best means by which to examine the prisoners’ dilemma and under such an analysis an egoist’s action does not prevent the maximisation of self-interest and is neither irrational nor incoherent.

§5.1.3.1 Gauthier – The Impossibility of Rational Egoism

Gauthier (1974) tries to advance game theory further and demonstrate the impossibility of rational egoism. Gauthier’s argument fails on several accounts but is still worth exploring as it is one of the best attempts to disprove egoism via a well developed application of game theory. Firstly, it must be pointed out that Gauthier does not deny that an agent could be a successful egoist. That is, an agent could successfully subscribe to egoism and simply never encounter a situation where the theory was unable to cope. Rather, Gauthier’s (1974, p. 441) argument is an attempt to show that “the conditions of egoism cannot be satisfied by any function defined over all possible situations in which one might act and specifying an action for each of those situations.” In short, situations occur where egoism fails to meet the functional requirement of providing an answer to the question ‘How should I act in this situation?’⁶³.

Gauthier (1974, p. 442) defines an egoist as “a person who on *every* occasion and in *every* respect acts to bring about as much as possible of what he values”⁶⁴. I have previously objected to such definitions; an egoist is one who gives preference to a particular moral theory, it does not obligate the agent to act in that manner on each and every occasion. Rather, egoism informs an agent as to which action is morally preferable on that specific occasion. The egoist may simply say, ‘Yes, morally the action I ought to take is Φ but I am simply not going to act morally on this occasion’. A better description of the egoist’s beliefs might be

⁶³ See §2.6

⁶⁴ My italics

that, 'an egoist believes that the morally preferable action on every occasion is the one that would bring about as much as possible of what the egoist values'. However, Gauthier's argument does not rely on the definition of egoism and in order to proceed all that is needed is agreement to something like 'an egoist believes that egoism can provide a justifiable answer to the question, how should I act in this specific context.' Gauthier's thesis is that egoism is unable to meet this requirement in *every* context and thus is an incomplete theory of action.

Gauthier (1974, pp 445-451) sets forth the following conditions that any principle of action must meet:

Condition 1: (Full definition) A principle of action must prescribe one and only one strategy for each agent in each situation to which it applies (p. 445).

This condition seems reasonable and is similar to the point raised by Medlin (§4.1). If a theory offers more than one course of action or strategy then that theory fails in its task of action guidance, and is incoherent. A theory which determines that an agent ought to adopt strategy A and also determines that an agent ought to adopt strategy B, where the two strategies are mutually exclusive, fails to determine which strategy the agent ought to follow and ultimately says nothing. The rational egoist must accept that if egoism is a theory of action guidance then it will produce a definitive answer to the question 'What strategy should I adopt in this specific scenario?'

Condition 2: (Equilibrium) The product of the strategies prescribed by a principle of action in any situation must be an equilibrium outcome. (p. 447)

Equilibrium points are part of a games theory strategy developed by J.F. Nash. A point of equilibrium is deemed to have been achieved when no player can benefit from changing their strategy while the strategy of the opponent remains unchanged. Consider the criminals' example; if both players had decided to stay quiet then either player could benefit by confessing, so long as their opponent's strategy remained unchanged and thus mutual cooperation (not confessing) is not an equilibrium point within the prisoners' dilemma. However, if both confessed then neither player can gain an advantage and, this strategy meets the condition of

equilibrium (at least when the game is viewed as a single event without the possibility of further repercussions). Two egoists, each knowing the other is an egoist, will act so as to attain an equilibrium point, as this is the best outcome they can rationally expect to achieve.

Condition 3: (Accessible Domination): The product of the strategies prescribed by a principle of action in any situation must not be an equilibrium outcome accessibly dominated by an outcome stable under reduction. (p. 452)

This condition is simply that if there are two or more equilibrium outcomes then the strategy that ought to be selected is the dominant one (i.e. where one player cannot make another move to regain advantage). Gauthier offers no proof of this condition, simply claiming that the condition is intuitively correct.

Condition 4: (Reduction) If a principle of action prescribes certain strategies to the persons in a given situation and if we reduce the situation by assuming that some of those persons follow the prescribed strategies, then the principle must prescribe the same strategies to the remaining persons in the situation so reduced. (p. 450)

By reduction, where a strategy of action for three or more persons results in more than one non-dominated equilibrium point, if one agent's action is taken as having been made (and is unchangeable) then equilibrium for the reduced set of agents must remain stable. The equilibrium point for three agents must remain the same for two agents, following the same strategy, once the decision of one of the agents has been fixed.

Assume the equilibrium for three agents is met if they all act in manner A or all act in manner B. We can reduce the scenario by fixing one of the agents' decisions at either A or B. If we then find that in the two-agent scenario either equilibrium A or B is accessibly dominated by the other then that option is unstable at the reduced level. Therefore, even though A and B are equilibrium points for the three agents, it is not the case that they can adopt either strategy; they must adopt the one that is stable at the reduced level.

According to Gauthier (1974, p. 454), if a theory of action is a complete theory of action then it must meet all four of the specified conditions. Gauthier claims that

egoism is unable to meet these conditions in specific contexts and therefore is only a partial theory of action (i.e. it may be possible to apply the theory but only because an agent is never presented with a specific type of dilemma).

The specific example Gauthier (1974, p.454) presents is one involving three people (A, B & C) where:

- (i) Each has exactly two pure strategies.
- (ii) A has a pure strategy a_1 and B has a pure strategy b_1 such that each agent is afforded at least as great a utility by this strategy as by his other pure strategy in all circumstances, and a greater utility unless the other agent uses his other strategy b_2 or a_2 and C uses his pure strategy c_1 .
- (iii) If A uses a_1 and B uses b_1 , C maximises his utility by using c_1 .
- (iv) If A uses a_2 and B uses b_2 , C maximises his utility by using c_2 .
- (v) If C uses c_1 then the utility for each of A and B is greater if A uses a_2 and B uses b_2 , than if A uses a_1 and B uses b_1 .

In such a situation the equilibrium point exists at $a_1b_1c_1$, where no agent is put at any disadvantage. However, if C's choice is fixed at c_1 then by reduction the equilibrium a_2b_2 dominates a_1b_1 and therefore (by condition 3) agents A and B must select a_2b_2 . However, prior to reduction agents A and B must select the equilibrium a_1b_1 . This is in violation of condition 4 that a theory must prescribe the same action both before and after reduction.

If Gauthier is correct, a serious theoretical problem has been outlined and egoism is reduced in status to a limited (partial) theory of action. However, a serious weakness in Gauthier's argument is his total reliance on Nash's theory of equilibrium. As Poundhouse (p. 99) notes, at times "Nash's equilibriums appear to be strictly irrational". Consider for example two rival manufacturing companies where each company initially sets its profit level at 32%. Each company can increase sales by decreasing profit levels and undercutting the other company. Each occurrence of undercutting is a move toward an accessibly dominated equilibrium. If conditions 2 and 3 hold then each company must reduce profit levels until they reach an equilibrium point that is not accessibly dominated. The only point at which this occurs is at a profit level of 0% where neither company can then improve its position by undercutting. However, while this position may be a stable equilibrium reducing profit levels to 0% appears anything but rational.

A more likely scenario is that the companies will agree on a higher level equilibrium, even though it 'could' be dominated, in order to avoid a series of *tit for tat*-like undercutting that could only result in the failure of both companies. Alternatively, one company could refrain from a final revenge if it believed other factors were involved (for example that the other company could not survive at the lower rates of return and it would eventually dominate the market)⁶⁵. Either way, we have reason to doubt that condition 2 (equilibrium) and condition 3 (accessible domination) hold fast in every context. If condition 3 is not absolute then the type of scenario Gauthier proposes poses no problem for egoism. The egoist need not change strategies based on the fixing of a third party decision. The egoist may overlook a dominating equilibrium, that is the result of a reduction, in order to avoid a revenge situation developing from abandoning the initial strategy. a2b2 may dominate once C has chosen c1. However, in order to avoid a revenge response from C, A and B's best strategy may lie in a1b1 despite any immediate advantage that deviating from this strategy might bring.

I am doubtful that *always defect, non-accessibly dominant equilibriums, or tit for tat* type strategies are sufficient for the analysis of actions outside of game theory. It is not possible to capture all of the contextual detail that is relevant to typical moral decision-making in the sort of scenario these theories can be applied to. However, assuming that these scenarios are a legitimate test of the capacities of a moral theory (regardless of whether they can actually occur), I have shown that it is not obvious that the egoist would adopt an *always defect* strategy or that an egoist must adopt a strategy that is a non-accessibly dominated equilibrium. Egoism is not shown to be incoherent by game theory analysis of *prisoners' dilemmas*.

I will now turn my attention to another attempt to prove egoism incoherent, namely, that any theory which makes reference to self-interest must be incoherent because the very concept of a self is incoherent.

⁶⁵ Also taking into account that each cut as the profit margin approached 0% is likely to have a diminishing effect on the increase in market share.

§5.2 The Nature of a Self

The self is a, perhaps *the*, central concept within egoism. If it can be shown that the concept of a self is incoherent then the concept of self-interest must also be incoherent and with it the doctrine of egoism. Critics of egoism who aim to show that it is based on an incoherent concept of the self, typically make one of two claims:

- (i) The self at any one time cannot be accurately demarcated.

If the boundaries of the self cannot be demarcated the correct beneficiary of an egoist act cannot be identified consistently.

- (ii) The notion of a self existing over time is incoherent.

This claim is similar to (i). However, where (i) mostly refers to the physical extension of the self, (ii) refers specifically to the extension of the self through time. The criticism is that any theory which purports to act for an agent's long-term interests is incoherent, for the simple reason that those long-term interests are the interests not of *myself* but of some *future* self; in effect, egoism is just a misguided form of altruism.

§5.2.1 Zemach's Self

For egoism to be a rational policy of action, the distinction between oneself and others (and thus between actions that benefit oneself and actions that benefit others), must be a factual one, not a mere conventional one. This, however, is not the case (Zemach, 1978, p. 154).

The central thesis of Zemach's criticism is that the way we divide up the world is a matter of convention, and while these conventions have been developed over time and, in the sciences at least, tend not to be completely arbitrary, neither are they factual. However, unlike the sciences and the so-called *laws* of nature, where the lines of demarcation are pragmatically sound, the lines that demarcate the self can become very vague and troublesome. I suggest the notion of 'I' that would be generally accepted (outside the philosophical community) would be something similar to:

I, the speaker, exist from the moment of my birth (perhaps slightly earlier) until the moment of my death.

However, it is quite clear that this is not the only criterion that could be used. For example, any of the following, (amongst others) might be used to demarcate the self:

Self-Awareness – The self might refer to a single person from the moment they gain self awareness until the moment they lose it; perhaps early childhood to senility.

Multiple Personality – It might be claimed that the self involves a single stream of awareness. In the case of multiple personality it might be claimed that there are several streams of consciousness existing within the same physical space and hence two or more ‘selves’ inhabiting the same physical space.

DNA structure – The self could be tied to a specific physical body. Then, in the case of exact clones either one self would inhabit two locations or two, identical yet separate, selves would exist. Further, despite the logical possibility of basing the self on DNA structure this concept appears somewhat disconnected from what we generally refer to as the self which tends to incorporate a sense of awareness.

Temporal Slices – The self may exist only for a specific period of time and bear only resemblance to past and future selves. (I will be discussing temporal slices in (§5.22) as they form the basis of the criticism that egoism, if it is based on the notion of a self existing over time, is incoherent).

Body parts, souls, and a host of other concepts might be introduced to demarcate the self, and with each of them their own set of problems. The commonly accepted division of the self into birth (or just before) until death is simply “a matter of pragmatic convenience ... it is not *wrong* to split up the world differently” (Zemach, 1978, p. 151). Any of the given definitional criteria could be used to denote the boundary of the self and “nothing in *nature* determines how far in time does the agent of action *x* extend before and after the time he performed *x*”.

If the self cannot be demarcated then what does it mean to act for my self-interest? If, by criterion *a* the self is *me** and by criterion *b* the self is *me*** then should I act so as to promote the interests of *me** or *me***, given that promoting the interests of one may harm the other? Zemach claims that an answer to such “pseudo-questions is logically impossible” (1978, p.454), the self cannot be objectively demarcated and therefore it is incoherent to speak of an action as

being in my self-interest. According to Zemach the best we can achieve is something along the lines of ‘the action x will benefit the person demarcated by criterion a ’. I contend that Zemach is wrong in stating that the self cannot be demarcated and that the mistake rests largely in assuming that the demarcation of the self, if it is to be demarcated, must be done via the establishment of objective criteria. However, before going into details I will briefly evaluate Zemach’s claim that egoism can be revised into a rational theory that does not require demarcating the self.

§5.2.1.1 Egoism Without the Self

Zemach attempts to formulate a theory of egoism that does not rely on demarcating the self and hence is immune to the criticism discussed above. Zemach determines the rightness of an action by application of the following principle:

For any action x , x is right iff it causes more benefit than harm. The amount of these [benefits and harms] is determined by adding the amounts of gratification and frustration caused by x to every y , weighted by the value of y to its agent Y and by the ‘distance’ between Y at the time of y and the agent of x , X , at the time of x . (Zemach, 1978, p.156)

An egoist (X) when performing an act (x) can determine the rightness of an act by evaluating the amount of benefit/harm done to everyone affected by x and the degree to which each individual (Y) affected by x has “Continuity, Contiguity, [and] Similarity” (Zemach, 1978, p.156) with X . In mathematical terms the utility of an act is calculated by multiplying the utility of an act x to each affected target by the similarity of that target to the originator of x .

For example:

I (now) as the son of A and B and having a certain set of values and considering an act x that will benefit me^* and harm me^{**} will act so as to benefit me^* if me^* is also the son of A and B and has a precisely similar set of values and me^{**} whom, while also the son of A and B , has a different set of values.

It is the similarity of me^* to the agent about to act which defines me^* as the preferable target of the act. Similarly, according to Zemach (1978, p.157) the egoist will be more committed to the well-being of dogs than microbes as “it is

probably easier to conceive of himself as a dog than a microbe”. A dog is more humanlike than a microbe and closer in kind to the agent than is the microbe. Therefore, when considering an act that will either benefit the dog and harm the microbe or benefit the microbe and harm the dog Zemach’s egoist ought to act so as to benefit the dog.

Zemach’s theory also has consequences that many would object to, for example:

I, as a male European, will be more committed to the well-being of males than females and to Europeans over those of other cultures.

However, it also has implications that appear decidedly un-egoistic. For example:

I_{now} as the son of A and B and having a certain set of values am considering an act x that will affect both me^* and me^{**} and will harm one while benefiting the other.

Me^* is a person with the exact genetic code as the agent about to act but many years in the future and (very probably) with a very different set of values (say the difference between a teenager and a middle aged person).

Me^{**} is my twin brother who has a markedly similar set of values.

Zemach’s egoist is morally obligated to act to benefit his twin brother and to harm the agent’s (genetic) future self, i.e., to act in un-egoistic fashion for the benefit of another over what most (non-philosophers) would have little difficulty in agreeing was the agent’s future self.

§5.2.1.2 Egoism, the Self, and Worldviews

I have some sympathy with Zemach’s position. Any attempt to objectively demarcate, that is, to declare the necessary and sufficient conditions of x , appears fraught with difficulties. As Wittgenstein noted, when considering the concept ‘game’ we find that it has no, or at least has very blurred, boundaries. If those boundaries are set too tight many things that we consider to be games are defined as being something other than games. If the boundary is set wide enough to include every possible case of a game (assuming such a definition was possible), then many things that are not games become defined as games.

While we can define the self within a specific context and for some specific purpose, all this does is provide us with a correct usage of a word within that very

specific context. For example, within the legal context the law must define a person if matters of ownership, personal rights, etc. are to be upheld. However, this does not provide an absolute definition of personhood, nor does it objectively demarcate a person outside of this context. All that is accomplished by a legal definition is that personhood refers to x when used within a legal context.

Perhaps Zemach is correct in that an objective demarcation of the self is impossible and all we have is convention. If this is the case then, given the variance in conventions (across cultures, beliefs, etc.), an objective identification of the correct target of an egoist's act is impossible and denoting x as being in X 's self-interest is simply incoherent. However, the egoist need not rely on an objective notion of the self, just as s/he need not rely on an objective notion of the Good. All the egoist needs do is consistently and, in a manner that fits within the egoist's worldview, demarcate what is meant when the egoist proclaims 'I'.

One possibility is to say that the egoist's claim is that 'I' (the egoist) ought to act in a manner to further the interests of the speaker of 'I', as defined by the speaker. The egoist is acting in the manner ascribed by egoism iff the egoist acts to preference whomever he recognises as forming a part of himself, over what he recognises as forming part of some other.

This however, is problematic, in that the speaker of 'I' could be mistaken in the identification of the speaker of 'I'. For example;

An egoist and three others are involved in an horrific car crash. The egoist regains consciousness, looks down and sees that (i) he has lost a limb and (ii) identifies a bleeding limb lying just within reach. The egoist believes the limb to be his own and so acts to slow his own bleeding and to save the limb. The limb is saved by the egoist's actions. However, it later eventuates that the limb actually belonged to one of the other passengers and that his own severed limb that was within reach of the egoist but obscured from view was irreparably damaged when rescue crews cut open the vehicle.

On evaluation it appears the egoist acted with inadvertent altruism in saving the limb of another, failed to act so as to promote his self-interest, and was incorrect in his identification of the boundaries of his self. However, the fact that the egoist

can be mistaken in some contexts is insufficient to dismiss a quasi-relativistic approach to the notion of a self-determined self. The problem in this scenario lies with the agent's lack of knowledge and not with the means of self-identification; self-identification is decisive under an idealised situation of perfect knowledge. The egoist acts in accordance with egoism so long as the egoist's actions are directed toward benefiting the egoist's self-determined self and in most cases the beneficiary of the act is the egoist's self-determined self.

The fact that the speaker can demarcate the extent of the speaker's self, relative to the speaker, is sufficient to enable an individual to act as an egoist. However, the speaker's ability to denote his own immediate physical extension might be insufficient as a complete basis for egoism. It needs to be considered whether the self-identification that has been applied to the physical extension of the self can legitimately be applied to temporal extension. Can an egoist, as the self-determined speaker of 'I', also legitimately claim that the speaker of I at this moment in time (I_{now}) and the speaker of I at some future time (I_{later}) are the same? If the answer to this question is no then egoism, or at least any egoist theory that incorporates claims about balancing long- versus short-term interests, is in effect nothing more than a misguided altruism. I_{now} in acting to benefit I_{later} acts so as to benefit some other (I_{later}) at the expense of the temporal self (I_{now}) and in direct contradiction of egoism. Thus there remain two questions to be answered in this chapter:

1. Can the self be demarcated as existing through time and, if so, to what extent?
2. If the answer to the first question is yes then can the hybrid nature of egoism (i.e., being agent relative yet temporally neutral) be justified?

§5.2.2 The self in time

I will not consider the claim that the self is nothing more than an instant and has no extension in time. Such a concept would rule out any form of morality (with the possible exception of some limited form of presentism). If we accept, as I assume most will, some form of self-based causality, some functionalist account that we have beliefs and desires, or that agents formulate reasons for actions that, if not directly causing, precede and lead to an agent acting, then the self must have some extension in time. However, while any form of causality that includes the

self requires temporal extension, such a self need not extend from birth until death.

According to Parfit selves are a series of mental states that are psychologically connected and while I_1 will likely have strong connection to I_2 , I_2 will have strong connection to I_3 , and I_n to I_{n+1} , I_{15} and I_{60} may have nothing in common. The teenager who takes out long-term life insurance may, according to Parfit, be so remote from the 60 year old who benefits so as to be completely detached and, in effect, different selves.

The position that a self may exist for only a segment of a biological lifespan and that a biological lifespan may consist of several different selves has its share of critics (see Persson (1985), Matthews (1998), Wolf (1986), Maitland (1989)) however my intention is not to attempt to disprove Parfit's thesis. For the moment I will accept that a self temporally expanded over a human lifetime and a self that consists of a life-stage are both possible, if problematic, explanations of the nature of a self. However, the notion that a teenager (I_{15}) and a 60 year old (I_{60}) while the same biological entity might be different selves goes against deeply held intuitions that the self does in fact exist across a (bodily) lifetime. As Robinson (2004) so aptly noted, "Earth carries about six billion uncontroversial cases of personhood". While these persons may all be mistaken about the nature of themselves and their continued existence from birth till death it goes firmly against the counsel of conservatism to throw this notion of a self away without very compelling evidence and, while the arguments of Parfit are philosophically interesting, they fail to offer the required compulsion to overturn the commonly accepted notion of a self through time.

Given that the egoist is more likely to subscribe to the common notion of the self than to Parfit's analysis, and that the way the egoist ought to act is determined from within the egoist's worldview (which incorporates the notion of a self across time) then, for the typical egoist, the notion of the demarcated self, both physically and temporally, is unproblematic.

§5.2.3 The Presentist Challenge

The egoist, in accepting the notion of a self extended through time and a premise that requires balancing future interests against present interests, is exposed to a theoretical challenge from the standpoint of a presentist; A presentist being someone who adopts the more minimalistic moral theory where interests are not only relative to the specific agent but also to the point in time when the agent holds those interests.

The presentist challenge to the egoist is that “if the utilitarian has to answer the question, ‘Why should I sacrifice my own happiness for the greater happiness of another?’ then it must be surely permissible to ask the egoist, ‘Why should I sacrifice a present pleasure for a greater one in the future?’” (Sidgwick, 1913, p. 418). Presentist theories, such as Parfit’s (1984) Present Aim Theory demonstrate “there is a coherent, more minimal alternative to rational egoism”. Of course, I am not contending that either the utilitarian or the egoist need answer this question as I have not claimed that egoism is *the* correct moral theory only that it is *a* rational moral theory. However, the challenge as formulated does provide an interesting lead into the question of whether moral theories must be pure (i.e., the terms used within the theory must be uniformly applied). If, as I have claimed, the egoist is justified in making the claim that egoism refers to a self extended over time, then the egoist needs to further explain why interests ought to be extended over time but ought not to be extended to include other agents.

A pure theory is one that has a consistent methodology. For example, theories of rational benevolence are both agent- and temporally-neutral, they involve doing good for others now or later. Likewise, presentism is consistent in that judgement of moral acts is relative both to the agent and to the moment in time. However, egoism is a hybrid theory agent relative (act for my good) but temporally neutral (act for my balanced future and current good). The hybrid nature of egoism has drawn criticism from Parfit (1984 p. 137-144), “If the egoist has no obligation to others then future interests cannot override present interests”, and Nagel (1986, pp. 88-100) if the egoist’s future self provides reason to act, then others also provide reason to act.

It would seem that if moral theories require parity (that is, moral theories must be either consistently relative or consistently neutral) then the egoist must either abandon the premise that egoistic action requires the balancing of long/short term interests and adopt a presentist position, or show that moral theories do not, in this instance, require parity. I contend that an egoist can justify the hybrid nature of the theory, for two reasons. Firstly, the hybrid nature of egoism reflects the hybrid nature of an agent. Given that the egoist can perceive of himself as a separate and individual agent extended over time, it makes sense for a moral theory to reflect this, and also to make a distinction between agents in space and agents over time. Secondly, as Brink (1992, p. 212) notes, “As long as agents are metaphysically distinct, interpersonal compensation is problematic; and as long as agents are temporally extended, diachronic, intrapersonal compensation is automatic”. Together with the popular notion of a self this view allows the egoist to deny the need for parity between different agents, and the same agents over time. In many circumstances acting for another requires a sacrifice on the part of an agent and, if Sidgwick is correct, it is unreasonable to demand that agents ought to make these sacrifices uncompensated. Such compensation may be by way of reciprocity (the expectation that the other will return acts in kind) or the personal satisfaction and good feeling an agent gets from acting in that particular manner. The difference for the egoist is that in acting for others there is no guarantee of compensation (unless the act makes the egoist feel good). However, when an egoist sacrifices a present good for a later good, provided it is accepted that personal identity does extend over time, the egoist’s compensation is automatic; I sacrifice and I am compensated for that sacrifice.

Physically demarcated and temporally extended selves are central to a conventional conceptual scheme. Given no compelling reason to renounce such a worldview there seems no good reason for denying that a theory of action, based on self-interest, should not be agent-relative and temporally neutral.

§5.2.4 Conclusion

My intention in this chapter was firstly to show that following the egoist doctrine does not prevent the egoist from attaining the egoist's goal. Maximisation of self-interest can be attained by adopting best strategies, that is, strategies designed to maximise self-interest in the long term even if initial risks must be accepted. The prisoners' dilemma only highlights a flaw in limiting consideration of interests to the present moment and, perhaps, highlights a substantial flaw in presentist theories.

Further, I argued that while it may be difficult, if not impossible, to objectively define the self the egoist can operate with his or her *own* concept of the self; which will likely match the conventional notion of a self. I have argued for what Brink (1992, p. 209) refers to as "a common-sense [notion of] personal identity over time". The egoist is justified in adopting a hybrid version of egoism (agent relativity and temporal neutrality) as this is the most accurate reflection of the hybrid nature of a metaphysically distinct and temporally extended self.

Chapter 6

Moral facts

§6.0 Is Egoism a Pernicious Doctrine?

The last of the criticisms outlined in chapter 1 (§1.2.7) was that egoism is an evil doctrine, and anything that is an evil doctrine is the antithesis of morality. Such lines of attack are uncommon in that it appears a moral judgment is being made (that egoism is an evil doctrine) prior to the conclusion of the argument that is supposed to determine the validity of egoism as a moral doctrine. If egoism is a standard of right and wrong then to pre-judge egoism as wrong (evil) is to beg the question. However, Rachels believes that this type of argument can be successfully formulated so it avoids begging the question and that showing egoism to be morally pernicious is “the most obvious objection to ethical egoism” (1974, p. 308).

Even if Rachels argument is successful it does not demonstrate that egoism is a flawed theory of action guidance in as much as it will still meet the minimalist definition as laid out in §2.6 and quite clearly does answer the question ‘How ought I to act?’. However, it is generally accepted that (a) morality is inextricably linked with some notion of the good life, and that (b) wickedness is contrary to such ideals. If we accept both (a) and (b) then any argument that conclusively shows that egoism compels, or at least provides strong reasons for, an agent to commit wicked acts must be taken seriously⁶⁶. What Rachels’ initial argument does demonstrate is a need to extend the minimalist definition and make a distinction between a theory of action guidance and a moral theory of action guidance. I contend that a theory of action guidance is one that consistently and coherently answers the question ‘How ought I to act?’ and that a moral theory of action guidance, in answering this question, does not compel an agent to act in a manner that is *clearly* morally wrong.

If Rachels is successful he will have presented a damning argument against egoism and shown that while egoism may be a theory of action guidance it is not a

⁶⁶ While both (a) and (b) can be denied there seems little point in doing so and to claim that morality is still the concept under discussion.

moral doctrine given that, in some contexts, it compels an agent to act in a morally pernicious manner. However, the question of whether specific acts can be recognised as clearly wrong must be addressed if Rachels' argument is to be accepted and the minimalist definition extended to include the distinction between theories of action guidance and moral theories of action guidance.

§6.1 Rachels' Argument

Rachels (1974, pp. 308-309) presents the following (true) case:

A friend of mine who lives in a small town in south Georgia told me about the following incident which occurred within the past year. The town is so small that there is only one doctor and he is, as one might expect, one of the town's more affluent citizens. One day the doctor was visited by a poor, uneducated black woman with a variety of minor complaints. A brief examination showed that she was suffering from malnutrition. The problem was that the woman did not have enough money to buy food for herself and several small children. She worked, whenever she could, as a cleaning-woman in the homes of the better-off people in the town, but she was able to earn only a few dollars a month in this way. All this was known to the doctor. After spending no more than five minutes with the woman, and having done nothing for her, the doctor told her the charge would be twenty-five dollars. The woman only had twelve dollars – this was, literally, all the money she had in the world – so the doctor took that.

Using the above case as contextual evidence and making an appeal to moral intuition Rachels (1974, p. 309) attempts to discredit egoism via a simple *modus tollens* argument:

- (1) If ethical egoism is correct, then the doctor did the right thing.
- (2) The doctor did not do the right thing.
- (3) Therefore, ethical egoism is not correct.

Correctly formulated *modus tollens* arguments tend to be uncontroversial and are generally considered to be valid. In this respect if someone is convinced that ethical egoism endorses the act as laid out in Rachels' scenario, and is also convinced that the act is wrong, then rationally that person must accept the conclusion that ethical egoism is a pernicious doctrine. However, I suggest that Rachels' intended purpose is not just to convince the casual reader that egoism is a pernicious doctrine but rather that the act outlined in the scenario *is* wrong and that ethical egoism *is* a pernicious doctrine. If this is the case then we need to do more than just examine if the argument is convincing, we must enquire into the

truth of the premises. If the premises of the presented argument are true then the conclusion must also be true and a decisive blow has been dealt to ethical egoism; it will have been demonstrated that egoism, in its consequentialist form, is not a moral doctrine.

As I noted in the introduction, the conclusions of such an argument are usually dismissed without much consideration, as it appears the argument is either viciously circular or question-begging. The act of taking the woman's last twelve dollars has been pre-judged as wrong and this claim (if true) in itself invalidates both the initial premise and egoism, and renders the conclusion of the argument redundant. If Rachels' criticism is to be taken seriously, and not dismissed out of hand as nothing more than a rhetorical ploy designed to discredit egoism, then the countercharge that the argument begs the question and that the second premise presupposes the conclusion must be addressed. Rachels' argument could be defended by appeal to either (noncircular) common morality, pure intuitions or by reference to moral facts.

In Chapter 7 I will examine in more depth the criticism that Rachels' argument begs the question and will argue that it cannot be defended by appeal to pure intuition or common morality. However, a more complex defense is available; namely that Rachels' thesis can be supported by an appeal to moral facts. If it can be shown that some acts are as a matter of fact wrong and that egoism in some contexts promotes such acts then Rachels will have demonstrated that egoism is a pernicious doctrine and therefore not a moral theory. It is to the possibility of moral facts that I will now turn my attention by engaging with Harman's (1977) anti-realist position.

§6.2 Harman's Thesis

Harman would reject the claim that Rachel's argument can be supported by an appeal to moral facts, arguing that we can have no such knowledge. Harman claims in his anti-realist thesis that observational judgments, such as 'X is wrong', provide no confirmation of moral facts and that moral facts have no effect on what you observe. If we have no way of confirming the existence of moral facts then

we cannot claim to have knowledge of such facts and without such knowledge it cannot be claimed that specific acts *are* pernicious.

The problem is not that moral observations cannot be used to confirm moral theories; if we accept that opinion can be formed via observations then it follows that moral observations may confirm moral theories (Harman, 1977, p. 5). Rather, the problem for those who would lay claim to the existence of moral facts is that it is unnecessary to make any assumptions about the existence of moral facts in order to explain observational judgments such as, 'X is wrong'. There is no justification for the assumption that moral observations are true propositions, or that they entail moral facts.

Harman notes a disanalogy between scientific and moral theories. While both scientific and moral observations are empirical in nature, scientific observations appear to play a role in the confirmation of scientific facts, whereas moral facts, if they exist, are unconfirmable by observation and appear irrelevant to the best explanation of moral judgments.

Harman (1977, pp. 6-9) outlines what he perceives to be the "important apparent difference" between scientific and moral cases by comparing two observational judgments. The first involves a scientific experiment and the formation of the conclusion 'There goes a proton'. The second involves the witnessing of a so-called immoral act, the burning of a cat, and the formulation of the conclusion 'That's wrong'. The difference between the observation of the two events and the formation of the observational judgments is as follows:

The truth of the proposition 'There goes a proton', as formulated by a scientist observing a vapor trail in a cloud chamber, is confirmed by:

- (i) The observation of a vapor trail in a cloud chamber, and
- (ii) The hypothesis that a proton passing through a cloud chamber will cause a vapor trail.

That a proton actually passed through the cloud chamber is the best explanation for the observation of the vapor trail which in turn explains the observational judgment 'There goes a proton'. Further, the observation of the vapor trail provides evidence for the truth of the proposition 'There goes a proton' and

provides verification for the hypothesis that when a proton passes through a cloud chamber it will cause a vapor trail. The scientist is justified in making assumptions about physical facts in explaining both the observation (the vapor trail) and the observational judgment ('There goes a proton').

However, according to Harman, this explanatory/evidential chain breaks down in moral cases. Explaining the observational judgment 'That's wrong', formulated upon seeing children pour gasoline on a cat and setting it alight, requires no postulation of moral facts. Hypotheses about moral facts are irrelevant to either explaining the children's act or the observational judgment. The best explanation of why an observer instantly formulates the proposition, or comes to a spontaneous belief (Lycan 1986, p. 97) 'That's wrong' can be completely determined in terms of the psychological states and moral sensibilities of the observer; no assumptions about moral facts are required⁶⁷.

According to Harman, (i) observation plays an important *and vital* role in the formulation and verification of scientific facts whereas (ii) observation plays no similar role, nor lends any empirical force to the formulation or verification of moral facts.

§6.3 Appeals to the Best Explanation and Evaluative Facts.

It is important to note at this point that Harman has not claimed that moral facts do not exist, only that moral facts are an unnecessary component in the explanatory chain from the witnessing of a cat being burned to the judgment 'That's wrong'. That is, the simpler explanation in terms of psychological states and moral sensibilities is better. In effect Harman applies Occam's razor and determines that in making an explanation we should not make more assumptions than the absolute minimum that are required⁶⁸. That Occam's razor, or some similar principle, ought to be applied is to appeal to an evaluative fact. It would be a mistake to assume that evaluative facts play no part in the formation of the observational judgment 'There goes a proton'. The chain from the observation of

⁶⁷ 'Completely determined' in the sense of what is required on top of the purely visual input.

⁶⁸ Occam's razor is attributed to William of Occam and is sometimes referred to as the principle of parsimony which underlies scientific modeling and theory building (F. Heylighen, 1997).

a vapor trail to the observational judgment ‘There goes a proton’ could be much more complex, perhaps involving a premise regarding God. Harman, in accepting the direct link between the observation of the vapor trail and the existence of the proton, makes an appeal to the best explanation and in doing so an appeal to evaluative facts.

The causal chain between direct observation and observational judgments is reliant on an appeal to the best explanation. However, that an explanation that appeals to the minimum number of assumptions is superior is not a given and while the principle appears to be solid it is ultimately a matter of preference. Yet this presents something of a paradox, for while it is logically impossible to determine a superior criterion by which to rank hypotheses (see §2.3), if the explanatory criterion is to be of any use at all then we must accept that some explanations really are better than others. For any given observation O there will be numerous competing hypotheses H_1, H_2, H_3 , that explain the observation. Noncircular selection of H_x as the best (most probable) explanation requires a criterion by which to rank hypotheses. Even the importance placed on observation requires an evaluative judgment, namely that observation really is a better form of verification than some other form of evidence.

Within the scientific paradigm it is usually accepted that qualities such as predictive success, repeatability, generality and simplicity form the backbone of the best explanation. Yet these are, quite obviously, not the only criteria that could be used and if all evaluative facts are denied then the selection of criteria for best explanations is itself nothing more than psychological preference. According to Sayre-McCord (1988, p. 278) we will not be saying that one explanation really is better than another only that we have societal or personal preferences regarding explanations. Unless we can legitimately claim that one explanation *is* better than another we cannot claim that the best explanation for observing the vapor trail *is* that a proton is passing through the cloud chamber. Harman’s explanatory chain is broken.⁶⁹

⁶⁹ Appeals to the best explanations and the concept of overriding criteria were discussed in §2.3

Harman's thesis is contradictory if it is dismissive of all evaluative facts. "What the explanatory criterion presupposes is that there are evaluative facts" (Sayre-McCord, 1988, p. 278), or at least there are some evaluative facts, those necessary for the selection of the 'best' explanation. These facts are not themselves directly, or indirectly, testable against the world but are a necessary presupposition if the explanatory chain from observation to observational judgment is to hold firm.

However, that evaluative facts are a necessary presupposition to explanatory chains does not prove that evaluative facts exist. It is equally viable that appeals to explanatory chains face the same fate as appeals to moral facts under Harman's analysis. Just as Harman denies moral facts on the grounds that moral judgments can be better explained in terms of psychological states, so we can deny evaluative facts on the grounds that explanatory judgments can be better explained by reference to psychological states and social sensibilities. The problem is that if we have no factual basis on which to judge between explanations then *any* explanation is an equally plausible candidate.

It might be objected I have been too hasty. Perhaps we do have general problems with explanations, but that is not to deny that there is a fundamental difference between morality and science; namely, that observation plays a significant part in the verification of scientific facts and no part in the verification of moral facts. Like Lycan (1986, p. 89) I accept that there are epistemological differences between science and morality but, as Lycan goes on to note, these differences also hold between science and "logic, mathematics, [and] linguistics". All knowledge requires evaluation and, if such evaluation is to be anything more than a societal preference, an appeal to evaluative facts. A strict reading of Harman's thesis sees, at the very least, "epistemological facts consigned to the same flames" as moral facts. Acceptance of Harman's thesis, logically extended, leads not merely to moral nihilism but to extreme nihilism, i.e. the denial of all values and ultimately all epistemological facts. Such extreme nihilism is a hole that is difficult, perhaps even logically impossible, to climb out of. However, while nihilism may be logically impenetrable it is questionable whether it is possible to actually hold such a stance and live in the world. I offer no argument against extreme nihilism, nor anything to try to sway those who hold this view. My resistance to extreme

nihilism is tooth and nail, not philosophical; it goes against too many previously held beliefs and offers no rewards or cognitive advantages.

In order to further this thesis it is at this point necessary to briefly inquire what is meant by the term 'fact'. In order to make such a determination it is necessary that I lay bare certain metaphysical assumptions. For the most part these assumptions are constructivist in nature and support a limited form of realism which, for want of a better name, I will refer to as demi-realism.

My first assumption is that extreme nihilism is false, but this assumption needs some qualification. If nihilism is simply a denial of any absolute knowledge of facts with regard to things-in-themselves then I agree with both nihilism and Kant; Noumenal reality is unknowable in principle. Our knowledge is of phenomena (via the senses) and while there may be an indeterminate number of ways the phenomenal data can be interpreted, and thus an indeterminate number of realities, this is not to say reality is whatever we make of it. The position I hold is constructivist and neo-Kantian; like Hung (1997, p. 449) I take constructivism to be more than the claim that "Reality is an illusion, [and that] there is no mind independent reality". I suggest that reality itself (even if unknowable), our social nature, and our initial assumptions about the world (even if this is nothing more than a general acceptance of sense data) constrict the reality we can construct. Acceptance that reality is socially constructed and theory-laden is also to accept that there are limits to what we can claim as fact. (For the remainder of this thesis I will use Reality_r to denote noumenal reality and Reality_s to denote our socially constructed reality).

Facts, be they scientific, social, or moral are not statements about Reality_r but rather about those phenomena that can be more coherently explained within Reality_s and figure in our best explanations of the world. I take a position similar to Quine (1970). It seems that the further something is from the data of sensation, and the more that needs postulating in order to derive an explanation of the observation, the less solid the conclusion we claim as a fact. It is easier to let go of the notion of sub-atomic particles (such as protons actually traveling through the cloud chamber) than it is to let go of the belief, based on an observation, that there

really is a vapor trail in the cloud chamber. We accept and treat as fact that the vapor trail exists because the best explanation of our observation of the vapor trail is that we really do see a vapor trail in the cloud chamber. We accept and treat as fact that a proton really is passing through the cloud chamber *only* because scientists offer a plausible explanation (protons, when passing through a cloud chamber, will cause a vapor trail) that does not conflict with our macroscopic (closer to observation) beliefs.

Harman also resists nihilism, at least in its extreme form, and looks for possible answers in reductionism and naturalism (1977, pp. 13-21) before pronouncing that moral facts are “not useful even in practice in our explanation of observations.” (Harman, 1977, p. 23). However, what Harman fails to fully consider is that, while moral facts may not figure in the explanatory chain from observation to observational judgment, moral facts might figure in explanations at a different level. As has already been argued, evaluative facts are unnecessary in giving explanations but they are necessary for appeals to the *best* explanation. Moral facts may fulfil a similar secondary, tertiary, or quadriary role, (as I will argue some mathematical postulates do) and be necessary to, or expand upon an explanation even if not directly part of the explanation itself. Lycan (1986, p. 82) similarly considers whether theories might be indirectly justified via explanatory virtues and reflective equilibrium. Prior to examining these claims I will briefly digress and consider what might constitute the criterion of a best explanation.

Best explanations are necessary if we are to make any social sense (as opposed to a purely private, unconfirmable interpretation) of reality. I have argued that best explanations are reliant on the admittance of evaluative facts to our ontology. However, I note that this has an element of circularity; for, while explanations are not just a matter of personal preference and good reasons can be given for the selection of evaluative criteria, ultimately these reasons rely on an appeal to best explanations and unless these reasons were selected by appeal to a new (superior) criterion then these criteria are selected via reference to the criteria they are selecting.

This circularity is unavoidable given acceptance of nihilism with regard to knowledge of Reality_r. We have no absolute facts to fall back on as a starting point and thus must accept an initial axiom (that determines the initial criterion of best explanation) if we are to make any progress. The axiom that I adopt, and now make explicit, is that making sense of Reality_s primarily involves increasing predictive success without sacrificing existing beliefs (conservatism) unless doing so offers sufficient and compensating cognitive advantage. We make sense out of reality in that our explanations not only rationalise past events but also allow us to predict what will happen in relatively similar situations, evaluate these predictions in terms of our general beliefs, and in turn extend our beliefs and understanding of Reality_s. The virtues (or values) of a best explanation that I will adopt, and appeal to throughout this thesis, are those of Quine & Ullian. According to Quine & Ullian (1978, p. 67), if we claim that hypothesis 'X' is the best explanation for 'Y' then hypothesis 'X' *must* meet the requirements of the virtues of hypotheses; Conservatism, Modesty, Simplicity, Generality, and Refutability. These virtues are for the most part self-explanatory, however, for the sake of clarity I have footnoted expansions on each of the five virtues.

Virtue #1: Conservatism

Conservatism is both the strongest and weakest of the virtues. Strongest in that it allows the quick elimination of explanations that diverge radically from current worldviews (for example that everything I do is controlled by an evil demon); weakest in that sometimes the view that radically differs from generally accepted views turns out to be the better explanation (for example, claims that the world was round at a time when the general worldview held that the earth was flat). In short, the virtue of conservatism states that the fewer conflicts a new explanation has with our previously held beliefs the better, as this limits the changes we need to make to our overall worldview.⁷⁰

⁷⁰Rationally, where hypotheses and previously held worldviews conflict we *must*, if we are to retain the hypotheses, reject the previously held beliefs. The overturning of (often deeply entrenched) beliefs is difficult and “the less rejection of prior beliefs required, the more plausible the hypothesis – other things being equal” (Quine & Ullian, 1978, p. 67). Where one or more hypotheses are available the hypothesis/explanation that conflicts with no prior beliefs usually prevails. We are not tempted by an explanation that upsets the applecart when a more conservative, non-conflicting explanation is available.

Virtue #2: Modesty

This consists of two similar criteria:

- The ‘modest’ explanation is the one that is logically simpler.
- The ‘modest’ explanation is the one that is more humdrum⁷¹.

Virtue #3: Simplicity (Occam’s razor)

Simplicity consists of two conjoined criteria:

- The fewer terms required in explanation, the better.
- The more the simple explanation encompasses, the better.⁷²

The force of conservatism is, initially, very weak. We form and discard beliefs with extreme regularity. Conservatism grows in strength the longer a belief survives and the more it fits within our developing worldview. To tear down these developed worldviews to incorporate some new radical theory is to take great risk. The incorporation of small changes avoids such risk; we can easily disregard the change, or continue to build on it without major disruption to our higher order beliefs.

Explaining how a magician performed a card trick (from a new and sealed deck), “The Counsel of Conservatism is ‘[He used] Sleight of Hand’” (Quine & Ullian, 1978, p. 67). Competing explanations such as “he did it by luck”, “he used a marked deck”, or “he has paranormal abilities” are generally dismissed on the grounds they conflict with the more strongly held beliefs respectively that “he would not have performed a trick with only a slim chance of success”, “he did not have access to the deck to mark it”, and that “the laws of physics and biology make paranormal abilities a doubtful concept”. This is not to say that the truth of the matter *is* that the magician used sleight of hand only that this is the best explanation of our observation. However, the truth of the matter may diverge from the accepted explanation but given our reliance on explanations one that is more consistent with current worldviews is (generally) preferable to one that calls for a revolutionary upheaval of our beliefs. Further there is danger in too easily accepting an explanation that radically differs and conflicts with current worldviews.

“The longer the leap, the more serious an angular error in direction. For a leap in the dark the likelihood of a happy landing is severely limited. Conservatism holds out the advantages of limited liability and a maximum of live options for each next move” (Quine & Ullian, 1978, p. 68)

To radically revise worldviews in order to accept that the magician’s trick was accomplished via paranormal abilities may, if the explanation turns out to be flawed, leave a worldview shattered and an agent with an incoherent view of reality.

⁷¹That an explanation be modest, under the first criterion, is that it can be explained by a single logic predicate rather than a conjunct of two or more predicates. Quine & Ullian (1977) suggest that given the two hypotheses, (a) the car in my driveway is my car, and (b) my car was towed away last night *and* someone parked an identical car in my driveway, the first hypotheses is preferable due to its logically simpler form. Hypothesis (a) consists of a single predicate while (b) relies on a conjunct.

Where two single predicate explanations are available the more humdrum explanation is preferable. In explaining why my car has a flat tyre; the ordinary single predicate explanation (c) I ran over a sharp object (causing a puncture) is preferable to the grander (d) someone shot out the tire (causing a puncture).

Virtue #4: Generality

The testability/plausibility of a hypothesis presupposes Virtue #4. Two tests can never be identical therefore the hypothesis must be generalisable to at least a second case. The wider the degree of generalisability the higher the degree of belief we can invest in the hypothesis.⁷³

Virtue #5: Refutability & Predictive Success

There must be some possible observation that would falsify the hypothesis or explanation.⁷⁴

⁷²Firstly, the fewer terms a new hypothesis has to refer to the better (the terms referred to may well be hypotheses in themselves). The strength of the new hypothesis is inversely proportional to the number of terms it is reliant upon. Secondly, the more a hypothesis encompasses the better. A hypothesis that explains the occurrences in multiple contexts is preferable to one that explains only a single event.

The truth, in a complex world, may be far from simple. However, “our steps toward the complicated truth can usually be laid out most dependably if the simplest hypothesis that is still tenable is chosen at each step. It has even been argued that this policy will lead us at least asymptotically toward a theory that is true.” (Quine & Ullian, 1977, p. 72).

⁷³“When a way is seen of gaining great generality with little loss of simplicity or great simplicity with no loss of generality, then conservatism and modesty give way to scientific revolution” (Quine & Ullian, 1977, p. 75). The virtues tend to become more powerful as they increase, but require far more careful application. Despite the disruption to our entire system of intertwined beliefs we had good reason to move away from our flat-worldviews and from Newtonian to Einsteinian physics as the eventual result, despite initial disruption, was a far more coherent view of Reality.

⁷⁴If no such event is possible then the hypothesis serves no actual purpose. A hypothesis that allows for any and every event neither predicts nor explains. Hypotheses involving Gods may take this form; it is impossible to refute a theory that some God causes all of our sensations as there is no possible state of affairs that would refute the hypothesis. Anything that we could claim was a counter example could just as easily be explained in terms of some God causing the perception that led to that conclusion. Similarly psychological egoism (§1.0) was dismissed because it fails to meet the virtue of refutability.

A hypothesis, together with background beliefs implies predictions (i.e. via induction, an explanation of a past event leads to a hypothesis that leads to predictions of future events). Should such future events be other than the hypothesis predicts then either the hypothesis or the background beliefs require adjusting. Albeit in most cases we would need more than a single exception to a hypothesis to justify its dismissal; two objects failing to react in the manner anticipated, given the gravitational forces in play, is not enough in itself to dismiss Newtonian theory. But, several such observations could be.

However, “[any hypothesis] *can* be held unrefuted no matter what, by making adjustments in other beliefs – though sometimes doing so requires madness... The degree to which a hypothesis partakes of Virtue 5 is measured by the cost of retaining the hypothesis in the face of imaginable events. The degree is measured by how dearly we cherish the previous beliefs that would have to be sacrificed to save the hypothesis. The greater the sacrifice, the more refutable the hypothesis” (Quine & Ullian, 1977, p. 79).

These criteria of explanation are accepted both generally and, more specifically, within the scientific community. Harman gives no details of the criteria he would refer to in determining the best explanation but given his analogy to scientific explanation it seems reasonable to adopt the five outlined virtues of best explanation.

§6.4 The Vapor Trail and the Best Explanation

If Harman's thesis is tenable then we can justifiably claim that the best explanation (in terms of the explanatory virtues) for the observational judgment 'There goes a proton' is that the vapor trail was caused by a proton. On the other hand, the best explanation for the observational judgment '[Burning the cat] is wrong' will *not* involve moral facts. Prior to turning my attention to moral facts it is necessary to enquire as to whether Harman is justified in claiming that the proton explanation is the best explanation of the observation of a vapor trail, and thereby satisfies the explanatory virtues.

It is not the case that a scientist, casually turning a corner, happens to see a vapor trail passing through a cloud chamber and instantly reaches the observational judgment 'There goes a proton'. The path from observation to observational judgment is far more complex. Cloud chamber experiments involve the firing of a proton through a cloud chamber so the scientist is already convinced that a proton is passing through the cloud chamber. The hypothesis being tested is that a proton will leave a vapor trail, as it passes through a cloud chamber. The actual observation is of a *reaction* in the cloud chamber (i.e. the trail). This observed reaction is both (a) confirmation of the hypothesis being tested, not a test for the presence of protons, and (b) best (causally) explained by reference to a proton passing through the chamber.

To use a bizarre example; The hypothesis that the earth was flat could have been held, unrefuted (even after the globe had been circumnavigated), by the addition of *ad hoc hypotheses*. Perhaps a God (or some mysterious force) immediately transported you to a new starting point as you reached the end of a flat earth. However, the cost of holding such a hypothesis and the need for more and more bizarre ad hoc hypotheses to support it would be far greater (at some point) than the cost of letting go of the previously held beliefs.

That we observe a vapor trail in a cloud chamber could be explained in many ways. It could, for example, be argued that protons do not exist, and that instead God causes the vapor trail each and every time the experiment is performed in order to test our faith. However, given that we are attempting to formulate a best explanation of our perception then, as Kornblith suggests, it seems reasonable to do so in a way that leads to “an improvement of our epistemic situation” (Walker, 2000, p. 80). Predictive success, such as predicting that a dropped object will fall to the ground, plays a big part in our making sense of Reality_s regardless of any actual causal connections. While it may be the truth that we are deceived and that our observation is the result of interference by an evil demon, the more humdrum, more modest theory that we actually did see a trail and that it actually was caused by a proton is preferable.

The observation of a vapor trail gives indirect evidence for a proton passing through the cloud chamber. The best explanation for the vapor trail, that a proton is passing through the cloud chamber, seems justified given the constraints of Reality_s and given what we value in appeals to the best explanation. Given these constraints, Harman is correct in the claim that scientific facts play an important role in the explanatory chain from observation to observational judgment.

§6.5 Moral Facts and the Best Explanation

Nelson (1999, p. 62) claims that Harman’s realist explanation, namely that the formation of the judgment ‘That’s wrong’ is better explained by reference to psychological states alone, fails to meet the criterion of conservatism.

Consider the following beliefs:

- (i) some things (be they actions or character traits) are actually wrong,
and
- (ii) in some cases there is close to universal agreement that some act is wrong or some character trait is bad.

Given (i) and (ii) persist despite widely diverging worldviews any explanation that dismisses all moral facts and appeals instead to purely psychological states calls for the dismissal of some very deeply held beliefs and goes against the counsel of conservatism. However, while conservatism is “necessary to get any reasonable epistemology of any subject off the ground” (Lycan, 1986, p. 85)

explanations are not to be judged by this single criterion. If conservatism alone were the only defining criterion progress would never be made; the firm, near universal belief that the 'earth is flat' would never have been overturned. The question is not; does an explanation in terms of psychological states meet the criterion of conservatism? Obviously it does not. Rather, does this explanation offer a better understanding of Reality, such that it warrants overriding the principle of conservatism?

An explanation of moral observations in terms of psychological states is certainly more modest and metaphysically simpler than one that requires the postulating of objective moral facts⁷⁵. Further, the hypothesis meets the criterion of refutability, in theory at least; we would have refutation for the theory if two agents in the exact same psychological state formulated different observational judgments on viewing an exactly similar act in an exactly similar context⁷⁶. However, even if we accept that Harman's thesis provides an explanation that is simpler, more modest, and meets (at least theoretically) the criteria of refutability, it remains doubtful that this provides sufficient justification to dismiss the counsel of conservatism. If psychological states are to prove a superior explanation they will need to do so on the grounds that the hypothesis not only meets all the other virtues bar conservatism but also has greater generalisability.

If the two theses (explanation by appeal to moral facts and explanation by way of psychological states alone) are both applicable only to the domain of moral explanations then neither is more generalisable and the counsel of conservatism recommends rejecting Harman's thesis. It is with regard to generalisability that Harman's thesis is stronger; reference to psychological states can be used to explain away *all* evaluative facts. However, in evoking the generalisability of the hypothesis it becomes viciously circular and self-destructive.

⁷⁵ However, as will be seen in the final section of this chapter, the hypothesis has some work to do if it is to explain animals carrying out what appear to be moral acts.

⁷⁶ The agents must be in the exact same psychological state as any psychological difference between the agents could be used to explain away any differences in observational judgments. Given that Harman appeals to scientific standards this would be the only test of refutability that meets those standards, the variable being tested (in this case psychological states) must be identical in the test case.

Harman's thesis is superior (a better explanation) in that it is simpler, more modest, and offers greater generalisability.

Harman's thesis can be generalised to any thesis that involves factual claims where those facts are not directly, or indirectly, supported by observation.

Evaluative facts are not directly, or indirectly, supported by observation.

If evaluative facts are dismissed (as nothing more than psychological preferences) then claims that one explanation is better than another are false.

If we cannot rank explanations then it cannot be true that Harman's thesis is superior.

Harman (1977, p. 13) does note that the discussion on moral nihilism *suggests* a dependence upon a similar premise:

We can have evidence for hypotheses of a certain sort only if such hypotheses sometimes help explain why we observe what we observe.

However, Harman thinks this assumption too strong and allows that we can have evidence for facts that are reducible to other facts that we have observational evidence for. Giving the example of social facts being reducible to facts about individuals for which we can obtain observational evidence leads Harman to consider if moral facts might be reducible (ethical naturalism) before dismissing such notions as problematic. I will not discuss Harman's less than convincing dismissal of ethical naturalism here as it is not necessary to the point at hand. Even if we allow as *facts* those hypotheses that are:

- (i) directly or indirectly confirmable by observation
- (ii) reducible to facts that meet criterion (i), or
- (iii) those that are necessarily appealed to in support of (i)

this will still be insufficient for Harman's case. Evaluative facts do not fit within any of these categories.

Evaluative facts are not reducible any more than mathematical facts are reducible to physical facts. If anything the reverse may be true. As quantum mechanics becomes more accepted, so does (within the scientific arena) the notion that our everyday experiences of time and space may be nothing more than illusion and that all we can be certain of is mathematical truths. Nor are evaluative facts

necessary to giving an explanation; their importance is that they are implicitly necessary to claims that hypothesis X represents the *best* explanation. If evaluative facts as distinct from mere psychological preferences are denied then claims that X is a better explanation have no factual basis. It seems that, at the very least, if Harman hopes to hold the structure of his thesis intact he will be forced to allow a fourth type of fact:

- (iv) those necessary to the structure of explanatory theory.

Harman must allow irreducible, non-observable facts, which, while not part of an explanation, are necessary to the formulation of a *best* explanation. My claim at this stage is only that the explanatory chain from observation to observational judgment is longer than Harman assumes and implicitly requires an appeal to evaluative facts that are, in themselves, not confirmable, directly or indirectly, by observation.

§6.6 Counterfactuals

Sturgeon (1988) criticises Harman's thesis by way of counterfactuals: Is it the case that if a direct observer were to view a cat being needlessly burnt alive, but burning the cat is not as a matter of fact wrong, the immediate observer will still form the same observational judgment 'That's wrong'?

Consider the following synopsis:

- (i) Unnecessarily burning a cat is objectively wrong
- (ii) You view the children carrying out what appears to be the unnecessary burning of a cat and form the observational judgment 'That's wrong'.

Sturgeon points out that when we consider counterfactuals we are not limited to varying only the moral facts. If the actions of the children aren't wrong, that is to say burning cats is still objectively wrong but what the children are doing isn't, then the children must be doing something other than burning a cat. Sturgeon's conclusion is that moral facts are relevant and play a role in what we observe (1988, pp. 249-250).

Sturgeon's defense appears to be directed at a minor strawman. It is not enough to show that moral facts are relevant to explanations of observations (Harman can concede that unconfirmable facts might exist), rather it must be shown that moral

facts figure in the *best* explanation of what we observe, otherwise they are explanatorily impotent. Further, it is not clear that Sturgeon has even accomplished enough to show that moral facts are relevant to explanations; that we believe moral facts exist undoubtedly has an effect on our observations but this leads only to the conclusion that our beliefs play a role in how we interpret what we observe.

Both Sturgeon and Harman lay out the moral case as a single event (observation), and single events lack verifiability. It may well be the case that Harman is correct and the reason the observer, in this isolated case, drew the conclusion ‘That’s wrong’ was due to the observer’s psychological state. However, if we are to be serious in our search for moral facts we need an experiment that is, at least in theory, duplicable. Hypotheses explain sets of data and not isolated events.

§6.7 Artificially Constructed Asymmetry

Much of the asymmetry in Harman’s analysis of the scientific and moral observations is artificial, that is to say the asymmetry is constructed, and disappears when cases that are actually analogous are used. When observing a vapor trail in a cloud chamber the actual observation is (i) of a reaction, and (ii) of something external to the observer. That is, an independent observer witnesses a reaction in a cloud chamber, explains why the reaction is occurring by reference to a hypothesis, and then formulates the observational judgment ‘There goes a proton’. The observational judgment is confirmable and comparable; other independent witnesses of the same experiment, if they formulate the same observational judgment, provide confirmation. Further confirmation can be obtained via observers of relatively similar experiments, again assuming that they reach the same observational judgment. It is this repeatability and comparability that leads to faith in the observations, the theories that explain the observations, and the truth of the observational judgment.

With the moral case, as Harman sets it up, it is difficult to locate the reaction (i.e. something analogous to the vapor trail) that the observer is attempting to explain. It might be that the fire is supposed to represent the reaction, which is explained by a hypothesis regarding flammable objects, but this only leads to the

observational judgment that the children set the cat on fire (or something similar) which is determinable by direct observation alone. It would seem that the only other reaction is internal to the observer; the observer notes a reaction in themselves. That is, the observer, in witnessing a cat being burnt, reacts and then forms the observational judgment 'That's wrong'. However, if this is the analogous reaction then we no longer have an independent witness as the observer is now part of the experiment. While the observer (O) can note that each time O views a similar act O denotes it with the token 'That's wrong' but in the strictest sense this observation lacks both the repeatability and the comparability of the scientific observation. O can never compare O's actual internal reactions, denoted by the token 'That's wrong', nor confirm that O is actually experiencing the same reaction to an earlier, similar, observation. Nor can O compare the actual internal reactions with others who claim to have had similar reactions and who denoted them by reference to the same linguistic token. 'That's wrong' could be being used to denote a range of differing internal reactions.

The asymmetry between the scientific and the moral results from Harman's use of disanalogous cases. In the scientific case an independent and objective observer formulates an observational judgment upon witnessing an external reaction, a reaction that is explained by reference to both the observation and a hypothesis. When an objective observer is introduced and care is taken to make sure all the relevant features are analogous we find that the moral and the scientific show a great deal of symmetry. In the moral case, to make an analogous comparison, we need an observer (to fill the role of the scientist), a medium (to replace the cloud chamber) in which to observe a reaction (the vapor trail), and an explanatory catalyst (the proton) that figures in a best explanation (hypothesis) of why the observation was made.

I contend that a more analogous comparison would be something similar to the following experiment:

In a controlled environment a person witnesses an act (such as the children setting the cat on fire), and an independent observer (of whom the person is unaware) then witnesses whether a reaction occurs (such as the direct observer exclaiming 'That's wrong').

We now have a (theoretically) repeatable experiment and one where multiple independent witnesses (both of this event and of exactly similar experiments) can compare the data on the reaction of the person directly witnessing the burning of the cat⁷⁷. If it turns out to be the case that the majority of direct observers react in a similar manner then we have an analogous case⁷⁸. Further, given a consistent reaction in a direct observer witnessing a cat being burnt (just as the vapor trail is a consistent reaction to a proton being fired through a cloud chamber) it seems reasonable to attempt to explain that reaction by appeal to a best explanation.

How most people would react to witnessing a cat being set alight is an unknown. However, I expect I am not making too great an assumption, for the purpose of this argument, in suggesting that most people viewing the cat being burnt by the children would have an adverse reaction. If you think my claim is too strong then the scenario can be replaced by a stronger one, perhaps an act of genocide, where you agree the reaction of the majority would be adverse. Just as the scientist attempts to explain why following act X a vapor trail is witnessed in a cloud chamber, the explanation we are now looking for is ‘Why do we (independent witnesses) observe an adverse reaction in most people when they observe a specific event (such as the cat being burnt)?’

Harman’s thesis, that the adverse reaction (the observational judgment ‘That’s wrong’) of a majority of persons to some specific scenario is a product of those observers’ psychological states, is still a contending explanation. However, I contend that at this stage the explanation begins to lose its inductive force. Assuming that the direct witnesses are from different cultural backgrounds and hold differing worldviews then it needs to be further explained why they reacted in a similar (adverse) manner despite the likelihood that the direct witnesses are in a variety of different psychological states. I imagine Harman’s reply would be that

⁷⁷ Obviously, for moral reasons, we can’t, nor would we want to, carry out such experiments. It might be possible to create an experiment where volunteers were shown pictures or videos of events and the reaction to these were observed. However, the experiment would be different in principle; at some level the volunteers are likely to doubt that what they are seeing really occurred and this is enough to affect the way the volunteer reacts. It is certainly enough that it would not meet the scientific standards we are trying to duplicate (i.e. the test introduces additional variables) and would render the experiment disanalogous.

⁷⁸ In almost all cases where a proton is fired through a gas chamber a vapor trail appears.

similar psychological characteristics exist throughout the group and that this is attributed to a matter of training or ethos. Harman's thesis at this level remains quite tenable; indeed the thesis would be difficult to falsify at any level. However, to prove Harman wrong is not my intention, rather the question is, does Harman's thesis provide the best explanation? That is, does it provide an explanation that is better than one formulated in terms of moral facts. It could be contended that "Harman's skeptical, antirealist explanation of moral observations is no better than competing realist, non-sceptical explanations" (Nelson, 1999, p. 62) and that Harman has provided insufficient reason to turn our backs on moral facts. However, I will argue an explanation formulated in terms of human flourishing is far more plausible (that is, that such an explanation meets more of the explanatory virtues) than one based on psychological states. Further, I will show that we have indirect observational support for moral facts in that they play a necessary supporting role in any theory of human flourishing (and that any discussion of flourishing will appeal to moral facts) in much the way that appeals to mathematical facts play a supporting role in scientific hypotheses.

§6.8 Mathematics

Harman (1977, pp. 9-10) claims that mathematics, while not confirmable by observation, is relevant to observation explanations. Although it is not necessary to invoke mathematics to expound the explanation that the vapor trail was the result of a proton, it is necessary to appeal to mathematical hypotheses in order to support the physical theory of protons. Since mathematics figures indirectly in the formulation of the hypothesis invoked as the best explanation we have "indirect observational evidence for mathematics" (Harman, 1977, p. 10).

While it is true that physicists rely on mathematics as they continually posit newer and smaller particles to accommodate "new observations without much loss of simplicity", they conveniently ignore that mathematicians have been employing the same trick, "multiplying entities, positing ever weirder species in order to simplify theory" such as, the acceptance of an 'unsolvable' Pi, negative numbers, and the postulating of "imaginaries to make exponentiation generally applicable" (Quine, 1970, p. 17). While the posits of mathematics and physics can be confirmed by indirect observation and appeals to explanations, there is no way of

selecting between the competing explanations at a quantum level. The queerness of facts at a quantum level and the general disagreement about how the world is best carved up at a subatomic level seem to go well beyond any objection, such as the one by Mackie (1977), that moral facts are queer.

However, that mathematical and scientific theories are problematic on the lowest level of explanation is not reason to dismiss these theories and “[t]o call a posit a posit is not to patronise it” (Quine, 1960, p. 22). Posits are necessary if we are to make any (scientific) sense of the world and as Quine goes on to say “[e]verything to which we concede existence is a posit from the standpoint of a description of the theory building process, and simultaneously real from the standpoint of the theory that is being built. Nor let us look down on the theory as make-believe; for we can do no better than occupy the standpoint of some theory or other, the best we can muster at the time”. Facts exist only as best explanations within a necessarily theory-laden Reality_s.

It is ironic that Harman dismisses the possibility of moral facts, yet implicitly accepts equally problematic mathematical and scientific posits as facts without a more than superficial questioning of their status. To dismiss moral facts on the grounds that they posit entities that (apparently) play no part in observation but to allow posited entities (protons) that are explained by posited mathematical entities is to apply a strange double standard. Mathematical posits are only provable internally within a theory-laden framework. Harman holds moral theory up for examination and claims it doesn’t meet scientific standards. However we might just as easily turn the tables. Scientific and mathematical theories are subject to rigorous standards within that discipline, but philosophy is equally, if not more rigorous, in its own way. Philosophers struggle to overcome counter-examples, find necessary and sufficient conditions, and create airtight theories that are true in all possible worlds. Such standards are not present in physics, where “inconvenient counter-examples no longer appear destructive of one’s position” (Rundle, 1993, p. 23) and are simply dismissed, to be explained later, without even consideration that the theory has been disproved.

Imaginary numbers (and many other theoretical posits) are not confirmable by either direct or indirect observation and play no direct part in explaining posits that are themselves indirectly observable. Imaginary numbers are necessary only for the internal coherence of mathematics. Yet it seems we must accept these posits if we are to have a consistent theory of mathematics, which is in turn necessary for making sense of scientific theory, and ultimately necessary in order to explain our observations.

The scientific scenario that Harman presents requires that we accept multiple levels of fact (levels of best explanations), ever increasing in distance from the initial observation. When we observe a vapor trail (or any event 'X'):

On an initial level, the best explanation for the observation of X is that X is actually happening. We see a vapor trail in the cloud chamber because there *is* a vapor trail in the cloud chamber.

On a secondary level, we formulate the observational judgment 'There goes a proton' and find verification for this from the initial observation together with a best explanation for the cause of the observation.

However, it must be noted that on this secondary level we have already appealed to an evaluative fact, namely, that observation supplies reliable data. This evaluative fact itself cannot be verified via observation.

On a tertiary level mathematics finds support in that "scientists typically appeal to mathematical principles" (Harman, 1977, p. 10) in providing an additional level of explanation; in this case explaining the theory of sub atomic particles. This is an appeal to the best explanation for the hypothesis that itself figures in the best explanation of the observation.

On this level we also have a second appeal to evaluative facts when we select which criteria count as better in the formulation of best explanations.

On a quadriary level, given Harman's *carte blanche* acceptance of mathematical and scientific posits we must accept a fourth level of facts; those facts that are determined purely by their necessity within a theory itself. Even if Harman had offered a deeper analysis of mathematics he would still have been forced to accept these posits as facts, or deny mathematics any internal consistency.

On the lowest level of mathematics and science, posits exist as facts that are very distant from any observation. However, all of these facts are ultimately related to

observation; they are all necessary to the formulation of the best explanation of an observation. If moral facts can be shown to be valid on any of these levels then we are as justified in referring to them as facts as we are to any other variety of fact at the equivalent level. It is at the tertiary and quadriary levels that I will argue moral facts are as important as any mathematical posits. It is as necessary to appeal to moral facts in explaining the principles of flourishing as it is to appeal to mathematics to explain some scientific principles. I will argue that mathematical and moral facts are highly analogous and that while both are “such remarkable human achievements that they naturally invite wonder as to how they are possible” (Lear, 1988, p. 93), they are both necessary within a theory laden Reality, and, that they each have as much claim to the status of ‘fact’ as do explanations that are closer to primary observations.

§6.9 Best explanation – One Last Time

The explanation we are looking for is one that best explains why:

We (independent, objective observers) witness an adverse reaction in an immediate observer (in most, sufficiently similar, cases) of a specific act ‘X’.

We are attempting to explain why in the majority of cases someone witnessing X has an adverse reaction. The act (X) that I suggest we consider is where young children pour gasoline on and set fire to a human baby. While X can easily be replaced by some other act I am suggesting that there are some acts that are close to universally condemned and it is these we should examine, not the more debatable act of the burning of an animal.

Harman’s hypothesis is that the adverse reaction we observe in someone viewing X is best explained by reference to psychological states and moral sensibilities and we have seen that this is severely problematic in as much as via generalisation it leads to the dismissal of all evaluative facts and, in doing so, undermines the claim that the hypothesis really is a better explanation. As I have already noted my aim is not to dismiss Harman’s hypothesis but to introduce another contender and argue that this new hypothesis is viable, confirmable, generalisable and has none of the problems of Harman’s hypothesis, namely:

[H₁] Acts detrimental to flourishing tend to cause adverse reactions in an immediate observer.

H₁ may sound like a psychological generalisation however it is ultimately my contention that explaining human flourishing relies upon an appeal to moral facts and it is the recognition of these facts that leads to an agent having an adverse reaction. It may be the case that recognition of moral facts on a subconscious level leads an agent to develop certain psychological states which in turn leads to observational judgments such as ‘That’s wrong’ and with such claims I have no real quarrel. Moral facts still figure in the explanation. It is just that they are further removed from the observation.

This hypothesis is testable against the world. If we (independent objective observers) witness an adverse reaction in an immediate observer who is witnessing an act detrimental to human flourishing we will have evidence for our hypothesis⁷⁹. Now, of course, it cannot be the case that every time someone witnesses an adverse reaction in an immediate observer the conclusion can be drawn that an act detrimental to human flourishing has just been witnessed, any more than a scientist can justifiably conclude ‘There goes a proton’ every time she sees a vapor trail. Rather, we would hypothesise something like;

[H₂] If, in a given context (Y), an act (X) would be detrimental to flourishing then we would expect, all things considered, that in most cases anyone directly observing X in context Y would experience an adverse reaction.

If, for example, we believed that an unnecessary act that caused intense suffering was detrimental to flourishing then, by hypothesis H₂, we would expect to observe an adverse reaction in any immediate observer witnessing such an act. So, in our modified Harman scenario, a woman observes a group of children burning a human baby and exclaims ‘That’s wrong’ (or exhibits some form of adverse reaction).

We (the objective observer) observe this adverse reaction in the woman.

On an initial level, the best explanation for us observing the woman having an adverse reaction is that she *is* having an adverse reaction.

On a secondary level, the observation of the adverse reaction together with the hypothesis is indirect evidence that she has witnessed something

⁷⁹ Further, if the expected reactions are not present this would provide evidence against the hypothesis.

detrimental to flourishing. Further, the observed reaction provides verification for the theory.

If the scientist is entitled to claim, as fact, that (a) there really is a vapor trail, and (b) that a proton has just passed through the cloud chamber, then we are as entitled to claim that (a₁) the woman really is having an adverse reaction, and (b₁) that an act detrimental to flourishing has occurred.

At this point I am making no explicit moral claims. Rather, I am claiming that in Reality_s there are certain acts that are detrimental to flourishing. No doubt the concept of flourishing needs considerable examination, and I will return to this later in the thesis. At this point I am merely suggesting that some things (in this case massive and unnecessary pain) are detrimental to flourishing and that this figures in the best explanation of why immediate observers have an adverse reaction to viewing certain acts.

It is on the tertiary level that moral facts find their initial foothold. Just as scientists necessarily appeal to mathematical principles to explain their theorems so is it necessary to appeal to moral principles in order to fully expound theories of flourishing. Flourishing could be denoted in many ways, from simplistic survival and reproduction to having all one's (perhaps rational, perhaps hedonistic) desires met. In order to fully understand what it is to flourish we need to appeal to evaluative (moral) criteria; namely, what it is to have a good life. My use of the term flourishing is somewhat crude at this time, I will return to a discussion of this concept in chapter 8.

At this stage I will offer only a brief example of a necessary evaluative (moral) criterion, drawing on the work of McIntyre (2001). To flourish we need to be able to make sense of the world and making sense of the world requires appealing to best explanations. Best explanation with regard to my flourishing requires self-knowledge as "self-knowledge is necessary, if I am to imagine realistically the alternative futures between which I must choose" (McIntyre, 2001, p. 95). True self knowledge requires an absence of self-deception. At the very least we must be honest and truthful with ourselves if we are to flourish. Honesty and truthfulness also has a social component, it is via confirmation that hypotheses are formed and

this requires that we can trust the information supplied to us by others. It is worth noting that the moral virtue of honesty is not just necessary to flourishing but is also an essential part of *any* best explanation; If a scientist cannot trust that the equipment was prepared as asked, that a proton really was fired into that chamber, and that others honestly report the results of similar experiments then the observational judgment ‘There goes a proton’ will never reach the status of being a best explanation.

I have spoken in terms of flourishing, rather than human flourishing throughout this chapter. My reason for doing so is that we can find observational support for the hypothesis “If in a given context (Y) an act (X) would be detrimental to flourishing then we would expect, all things considered, that in most cases anyone directly observing X in context Y would experience an adverse reaction” as well as indirect support for moral facts from outside the human sphere.

Stingl (2000, pp. 247-250) outlines an interesting scenario regarding chimps⁸⁰ and the distribution of food. Given enough food chimps will share the food amongst the group. If any single chimp tries to take more than their share then the dominant male will discourage this act by dealing out punishment. Moreover, this punishment itself is judged by the females of the group and if the dominant male deals out too much punishment he will be chastised by the females of the group. It appears that when one chimp attempts to gain an unfair advantage, depriving another of sufficient food (an act that goes against the flourishing of the chimps), the dominant male has an adverse reaction and acts in such a way as to rectify the situation. It further appears that if the dominant male delivers too great a punishment to the offending chimp (an act that goes against the flourishing of the individual chimp and possibly causes distress in the group) the females act to remedy the situation. Further, these actions occur only when there is roughly sufficient food for the group. If there is an abundance of food chimps are not punished for taking more than they need. Likewise, during scarcity chimps are not punished for not sharing even though harm is done to those that go hungry (we might assume that when food is in extremely short supply, equally sharing the

⁸⁰ I have used the colloquialism ‘chimp’ as this is the term Stingl uses.

available food puts everyone at risk). Stingl notes that not only must the dominant ape be able to recognise when harms are justified but also the level of punishment that is warranted. Further, the female group members “can tell when a certain level of harm is justifiably imposed on another chimp and when it is not” (Stingl, 2000, p. 247).

Harman’s explanation of this scenario would deny that the chimps were acting because of any actual rights and wrongs. Rather, the actions were purely psychologically driven. That some acts are harmful to the chimp group is (for Harman) an objective fact, but that certain punishments are justified “only appears to be objectively true, to the deluded chimps, insofar as they are psychologically programmed to mistake appearances for reality in precisely this sort of way” (Stingl, 2000, p. 248). This explanation is of course a possibility but it seems to involve a somewhat complicated story given we are unlikely to ascribe to chimps the rational capacity to develop “moral sensibilities” and debate moral beliefs. A better explanation, given the appearance that the chimps’ actions are both justified and applied within specific contexts, is that acts that affect flourishing give rise to attitudes of approval or disapproval and that the chimps have evolved to a point where they recognise these rudimentary moral facts (even if they don’t recognise them *as* moral facts) and are compelled to act by them. We gain greater predictive success and gain far more understanding of the chimp society via an explanation in terms of flourishing and through postulating moral (virtue) facts than we do through ascribing their actions purely to psychological states.

§6.10 Conclusion

If Harman’s thesis is only that moral facts are impotent in the formulation of best explanations of observations, then the thesis is untenable. Firstly; the asymmetry between scientific and moral observation is mostly of Harman’s own making. Secondly; Harman’s realist thesis is better only if it is more generalisable than the anti-realist explanation. However, in its more generalisable form it will also defeat evaluative facts and in doing so deny any possibility of an appeal to the best explanation.

A stronger reading of Harman, despite his denial of such a conclusion, leads not to nihilism (moderate or otherwise) with regard to moral facts, but to nihilism with regard to all epistemic facts. I have not argued against this position. From a neo-Kantian perspective Reality_r is, in principle, unknowable. Yet this does not mean we have no claims to factual knowledge, rather those entities which we refer to as facts *are* the best explanations of specific phenomena experienced within a socially constructed reality (Reality_s).

It is here I strike bedrock and claim, without argument, that the best explanations are those that increase predictive success without sacrificing existing beliefs (conservatism) unless doing so offers sufficient cognitive advantage. These best explanations are facts within Reality_s and these facts allow us to make sense of both the world and our lives. Further, it seems upon acceptance of the initial axiom that *there is no cognitive advantage in not postulating moral facts*. Rather, the opposite is true. We gain great advantage in accepting that explaining the concept of flourishing requires appeal to moral facts in the same manner that explanation of science requires an appeal to mathematics.

Ultimately mathematical, moral and epistemic facts are nothing more than constructs within Reality_s. However, that something is a posit is not to deny it status as fact; within Reality_s and divorced from theory everything is a posit. Facts in this context can only be appeals to the best explanation of Reality_s and it is only by appeal to moral (virtue) facts that a best explanation of what it is to flourish can be formulated.

This chapter began with a consideration of Rachels' argument that egoism is a morally pernicious doctrine. Escaping the criticism that Rachel's thesis begs the question requires an appeal to moral facts such that the claim some acts are pernicious can be substantiated. If facts are seen as best explanations, as I have argued they must be, then moral claims gain the status of facts and can be used in support of Harman's thesis. In the following chapter I will begin afresh with the question "Is Egoism a morally pernicious doctrine?"

Chapter 7

Consequentialism, Egoism, and Pernicious Doctrines

§7.0 Introduction

As discussed in the introduction to Chapter 6, Rachels attempts to prove that egoism is a pernicious doctrine by showing that adherence to the theory compels an agent, in some contexts, to commit wicked acts. Rachels tells the true story of a doctor who took a poor, uneducated, and malnourished black woman's last twelve dollars after giving her nothing more than a superficial examination, leaving the woman without means to feed herself or her small children. The doctor, a selfish man who will likely suffer no pangs of regret from taking the woman's last twelve dollars, meets a minor interest by increasing his personal wealth. Rachels contends that the doctor acted rightly according to egoism but that the doctor's act in taking the woman's last twelve dollars was clearly wrong, thus egoism requires the doctor to commit an evil act and is a pernicious doctrine.

The argument, in its simplest form, is:

- (1) If ethical egoism is correct, then the doctor did the right thing.
- (2) The doctor did not do the right thing.
- (3) Therefore, ethical egoism is not correct.

In §6.1 I noted that such arguments are usually dismissed as they beg the question. The second premise predefines right and wrong and in effect predefines egoism as a pernicious doctrine. I will argue that the only successful defense that can be offered against this criticism is by way of an appeal to moral facts. However, prior to outlining this defense I will briefly outline other possible responses to the criticism and the reasons why these defenses are unsuccessful.

§7.1 Question Begging

An argument is assumed to have begged the question if the conclusion is smuggled into the argument as an implicit premise, or if one of the premises is a disguised form of the conclusion. The criticism that an argument is circular or begs the question can be formulated in many ways. However, I will focus my attention only on the stronger claim that Rachels' second premise presupposes the falsity of egoism:

- (a) Assuming the first premise is true; if ethical egoism is correct then the second premise must be false.
- (b) Rachels claims the second premise is true.
- (c) If the second premise is true then the conclusion that ‘the doctor did not do the right thing’ must be the conclusion of a moral argument from an ethical perspective other than egoism. That is, some non-egoistic moral principle must have been applied by which ‘taking the woman’s last twelve dollars’ was judged as wrong.
- (d) [by c] Rachels’ argument is exposed as:
 - (1) If ethical egoism is correct, then the doctor did the right thing.
 - (2) According to *some other moral theory or principle, or by moral intuition alone* the doctor did not do the right thing.
 - (3) Because premise (2) *is* true it (question begging) is the case that ethical egoism is not correct.

The main force of the criticism is that Rachels has assumed, without argument, that moral intuition or some non-egoistic moral theory or principle, that predefines the act recommended by egoism as wrong, *is* correct. It is trivially true that according to *some other moral theory or principle* egoism is incorrect but this is a far cry from proving the conclusion that egoism is incorrect.

§7.1.1 Question Begging – Rachels’ Defense

Rachels addresses this counter argument and claims that the second premise can stand alone. That is, premise two does not rely on some hidden principle or theory. As Rachels goes on to claim, the wrongness of taking unnecessary advantage of a helpless woman is intuitively recognizable independently of, and prior to consideration of, any moral theory or principle. Rachels would have us accept that the wrongness of the doctor’s act is clearly recognizable as a moral fact and that the wrongness of the act is “more certain than any mere theory could be: it is one of the fixed, constant points against which proposed theories may be tested” (1974, p. 313).

Rachels’ claim, that the act of taking a helpless woman’s last twelve dollars is recognizable as wrong by way of moral intuition alone, does not imply that moral intuition is foolproof nor that we can always identify right and wrong by way of

moral intuition alone. The claim is the simpler one that certain extreme acts (for example torturing an innocent for fun or, in this case, taking unnecessary advantage of a helpless woman) *can* be objectively identified as wrong and *will* be identified as wrong by the majority of individual moral agents. The implicit generalised claim on which Rachels' argument depends is:

We (presumably morally sensitive people) will recognise *intuitively* that certain specific acts (such as taking unnecessary advantage of a helpless woman) are objectively wrong with more certainty than we know the reasons why such acts are wrong.

That the doctor, in taking the twelve dollars, did not do the right thing is, according to Rachels, recognizable as wrong with intuitive moral certainty. The egoist, in claiming the doctor did the right thing, is mistaken (and, presumably morally insensitive) and has simply failed to recognise the wrongness of the act⁸¹.

Rachels offers no argument to support the claim that the majority of persons would make the moral judgment that the doctor acted wrongly in the detailed scenario and that egoists (in holding on to the claim that the doctor did the right thing by taking the woman's last twelve dollars) represent a small minority of the moral community. Nor, to be fair, does he really need to.

All Rachels needs us to accept is:

1. There is some act (X) where the moral intuition that (X) is wrong will be the intuition held by the majority of the moral community such that it is reasonable to conclude that (X) *is* wrong;⁸²
and
2. There exists a situation where the act (X) will be the act promoted by egoism.

It is the principle that some acts can be intuitively recognised as wrong and would be recognised as wrong by the majority of the moral community, not the specifics of the presented scenario that is central to Rachels' claim. If such a position stands up under scrutiny then Rachels' second premise 'The doctor did not do the right

⁸¹ It could be that the egoist recognises the act as wrong but chooses to follow the egoist doctrine anyway. However, it would seem such an egoist has already given up on the claim that egoism is a moral doctrine.

⁸² This is not an appeal to majority belief, or a utilitarian calculation. Rachels is claiming only that the majority will recognise an act as being wrong without reference to any belief set or ethical theory.

thing' is, as Rachels himself notes, recognizable independently of, and prior to consideration of, any moral theory.

§7.1.2 Question Begging at a Deeper Level

Rachels' defense against the criticism that his argument begs the question relies on the claim that agents can experience pure intuitions. It is undoubtedly true that we can have moral intuition without knowledge of the structure of any specific moral theory – children formulate the notion that someone has acted wrongly toward them long before they gain knowledge of why the act might have been wrong. However, it is not so obvious that these intuitions are pure. While someone may be unaware of the specific moral structure of principles it may still be the case that learnt, or indoctrinated, moral principles are the ultimate basis of moral intuitions.

As Hare (1986, p. 165) notes “[principles], if they are accepted sufficiently long and unquestioningly, come to have the force of intuition”. A child indoctrinated with a specific principle is likely to judge something that goes against that principle as wrong without any further consideration, and possibly without any conscious consideration of the indoctrinated principle. That Rachels and others intuitively find the doctor's act reprehensible may simply be a result of having accepted the guiding principle that ‘taking unnecessary advantage of a helpless person is wrong’ to such a degree that it now appears unquestionable and even the consideration of doing otherwise is, to them, abhorrent. If this is the case then Rachels' argument still begs the question. Rachels' second premise:

(2) [we can recognise by way of intuition alone that] the doctor did not do the right thing.

is really nothing more than the claim:

(2) [we can recognise by way of deeply held moral principles that have the force of intuition that] the doctor did not do the right thing.

Ethical egoism is being dismissed, not by reference to some presupposed moral theory, but on the basis that it does not conform to some deeply indoctrinated (non-egoistic) moral principle. Many put little faith in the process of reciprocal

illumination and decry it as an error to rely on intuition over theory.⁸³ However, intuitionism has its supporters. Though not specifically in a moral context, Kripke (1980, p. 42) claims “Of course, some philosophers think that something’s having intuitive content is very inconclusive evidence in favor of it. I think it is very heavy evidence in favor of anything, myself. I really don’t know, in a way, what more conclusive evidence one can have about anything, ultimately speaking”. Likewise Nelson (1999), while not offering anything new by way of argument for intuitionism, puts up a strong defense against Kantian and Aristotelian criticisms. I cannot possibly hope to do justice to, nor settle the debate regarding the purity of moral intuitions here. Rather, in order to make progress, I will for the time being work under the assumption that pure intuitions are at least logically possible and grapple with the question ‘why ought we take such intuitions as being indicative of a moral fact?’

§7.1.3 Question Begging – The Power of Intuition

The practice of what Sober (2001, p. 432) refers to as reciprocal illumination is common in moral philosophy. It is a logical truth that if in a given context our moral intuitions differ from what moral theory advises then, if we are to be consistent, either the theory must be disregarded (or modified), or we must go against our intuitions. However, it is a considerable leap from accepting that we have pure (uninfluenced) moral intuitions to assigning them any form of factual status. Likewise, what might be referred to as common morality where the majority views an act as wrong, intuitively or otherwise, does not provide proof that such an act is, as a matter of fact, wrong.

If moral facts are intuitively recognizable then we must question why Mary Wollstonecroft’s intuitions were those of a minority, and why her claims that women ought to have equal rights, were so “widely regarded as absurd... [and why] ...they were satirised in an anonymous publication entitled *A Vindication of the Rights of Brutes*⁸⁴” (Singer, 1976, p. 33). It could of course be countered that the majority did intuitively recognise that women ought to have rights but that

⁸³ Nelson (1999) cites Brandt, Hare, and Singer as regarding intuitionism as the “fashionable new error in ethics”.

⁸⁴ The author is now known to be Thomas Taylor.

these intuitions were suppressed for political reasons. However, we need not rely on this single example. The briefest look back through moral history shows the naiveté in claiming the moral intuitions of the majority represent (realist) moral facts.

The current majority view is that slavery (where slavery means the ownership of another human) is morally wrong. Yet, the ancient Greeks (and many others throughout history) appear to have had no doubt that slavery was morally acceptable. Few ancient Greeks would have recognised a premise suggesting that enslaving other humans was factually wrong or that such a claim was clearly an intuitively recognizable moral constant. It appears that at different points in history moral intuitions (or at least the moral intuitions of the majority) differed:

- (a) The ancient Greeks did not intuitively recognise slavery as wrong.
- (b) Most people in the modernised world do intuitively recognise slavery as wrong.

That moral intuition varies across time can be explained by:

- (i) The intuitions (a) and (b) about slavery are correct in both cases. Morality is temporally relative.
or
- (ii) Pure moral intuitions, where they seem certain and gain some credence in that the majority of others within the moral community share the intuition, can be mistaken.

§7.1.3.1 Temporally Relative Moral Intuitions

I will discuss the possibility that morality is temporally relative, that is that the morally correct action in a specific context is relative to the point in time at which the scenario occurs. More specifically I will consider the possibility that:

The intuitions (a) and (b) about slavery are correct in both cases. Morality is temporally relative.

While this would account for differing moral intuitions across time it presents a new set of problems and considerably weakens the conclusion of Rachels' argument. If it is accepted that moral intuitions are temporally relative then all that can ever be claimed is that 'X is intuitively wrong at this particular instant' and while the intuition may also hold for the next instant, and perhaps for a considerable time, there is no reason why this must be the case. Further, Rachels

is unlikely to want to adopt this position as it means premise (2) must be replaced with,

(2) The doctor did not do the right thing *relative* to this specific point in time.

and the conclusion of the argument is no longer that ethical egoism is incorrect, only that:

(3) Ethical egoism is incorrect at this point in time.

Temporally relative intuitions are unsuitable for reciprocal illumination. The correctness of moral theories would come and go with the passing of time; one moment correct, one moment flawed⁸⁵. If temporal relativity is accepted Rachels is forced to concede that, should moral intuitions change, his argument against egoism will lose all its inductive force. Egoism might be an incorrect theory today but it could be *the* correct moral theory tomorrow.

Temporally relative intuitions, while logically possible, are also an unsatisfactory answer. If universalisability is a criterion of morality within time then consistency demands it must also apply across time. This leads to conflicting intuitions, the Greeks' intuition that slavery is morally acceptable is applicable both then and now and the current (common) intuition that slavery is wrong is applicable both now and then.

Nothing I have said disproves the notion of temporally relative intuitions. However, if moral intuitions are morally relative then they are of no use in strengthening Rachels' argument.

§7.1.3.2 Moral Intuitions and Certainty

If moral intuitions are not temporally relative then it must be the case that:

- (iii) Pure moral intuitions, where they seem certain and gain some credence in that the intuition is shared by the majority of others within the moral community, can be mistaken.

⁸⁵ We might expect reasonably slow change but this need not be the case. Remarkable events that change our view of the world overnight might also lead to sudden changes in our intuition about right and wrong.

Either the intuitions of the ancient Greeks or those of the modern world are mistaken. It cannot be that *moral agents know with certainty* that slavery is morally acceptable and other *moral agents know with the same degree of certainty* that slavery is morally wrong. If it is conceded that moral intuitions are fallible (even where such intuitions coincide with the majority of others' intuitions) then Rachels' defense of premise two fails. We cannot be certain, by means of intuition alone that, as a matter of moral fact, the doctor did the wrong thing. Further, if moral intuitions that seem both obvious and certain can be false then it seems odd to ever assign them preference over moral theory.

However, if moral intuitions are not infallible then they cannot be appealed to by Rachels in order to avoid the criticism of question begging. Rachels' second premise (that the doctor, in taking unnecessary advantage of a helpless woman, by taking her last twelve dollars did *not* do the right thing) cannot be the conclusion of a moral theory or principle and, as I have shown, is unsuccessful in supporting Rachels' conclusion if it is based on moral intuition alone. However, I contend that a stronger version of Rachels' argument can be formulated by introducing the notion of moral facts.

§7.1.4 Question Begging – Moral Facts

In Chapter 6 I developed the thesis that all facts are constructs (best explanations of phenomena within a socially constructed reality) and that moral concepts have a factual status that is at least the equivalent of that of abstract mathematical posits. Where moral condemnation of a specific act approaches unison we have strong verification for moral hypotheses.⁸⁶ The type of hypothesis I explicated in Chapter 6 involved the notion of flourishing and I further contended that a full account of what it is to flourish requires an appeal to moral facts in the same way that the explanation of a simple mathematical procedure (square roots) requires appeal to a mathematical posit (imaginary numbers). Accepting that moral concepts, even if ultimately constructs, are a necessary part of the explanation of

⁸⁶ We do not need 100% unison of agreement any more than science expects a 100% success rate – though this is an ideal things sometimes go wrong in any experiment.

human flourishing⁸⁷ allows for the considerable strengthening of Rachels' initial argument, and for the argument to be formulated without begging the question.

Before continuing I will make clear how this position (regarding moral facts) is incorporated with the doctrine of egoism as developed in the initial chapters. My claim is that Good is relative both to an agent and that agent's specific worldview. The values and beliefs that an agent holds, within that agent's socially constructed reality (Reality_s), lead to certain actions being labeled as right or wrong. However, I am not claiming that any and every imaginable view of the world could be held by an agent. Reality_r (even if in principle unknowable), social interaction, and the nature of being human all constrain how we view the world (Reality_s). My claim, and the argument that will be further developed in Part 2 is that:

While denoting good as relative to the egoist and his/her worldview, the egoist must also admit, given the limitations placed on Reality_s, that some actions (as a matter of best explanation) really are for the benefit or detriment of human flourishing, and that the explanation of why these acts are perceived to be right or wrong requires an appeal to moral facts.

Returning to Rachels' case, my contention is that the reason most people experience an adverse reaction to hearing the story of the doctor taking the woman's last twelve dollars is that the act is detrimental to flourishing⁸⁸. Further, the general intuition that the doctor acted wrongly is an indication (Virtue #1: Conservatism) that the doctor did in fact act wrongly. These intuitions could be explained away by Harman's antirealist explanation with regard to psychological states and, despite the problems I have highlighted with this method of explanation (§6.2), this is still a potential explanation. However, my claim is that:

- (i) the doctor did something detrimental to human flourishing;
- and*
- (ii) in fully expounding why the act was detrimental to flourishing it will be necessary to make an appeal to moral facts;
- and further that*
- (iii) an explanation in term of flourishing and moral facts is a better explanation and does not result in the nihilistic position that an appeal to psychological states leads to.

⁸⁷ What flourishing is and who can flourish has not yet been discussed. For now, I am working with a general assumption that some states of affairs lead to human flourishing and others do not. I will return to the concept of flourishing and a more in-depth analysis in Part 2 of this thesis.

⁸⁸ Specifically to the woman and more generally to human flourishing in that we believe doctors ought to act in a specific manner.

Incorporating the above claims, Rachels' argument can be expanded to:

- (1) If ethical egoism is correct, then the doctor (in taking the woman's last twelve dollars) did the right thing.
- (2.i) Most people would have an adverse reaction upon witnessing a scenario whereby a doctor took unnecessary advantage of a helpless woman by taking her last twelve dollars.
- (2.ii) The best explanation for why the majority had an adverse reaction to the doctor's act, given no other overriding explanation, is that the act is detrimental to the woman's (and perhaps the doctor's) flourishing and that most people recognise the wrongness (in this case) in acting to the detriment of the woman's flourishing.
- (2.iii) [by 2.i/2.ii] It is a moral fact within Reality_s that taking unnecessary advantage of a helpless woman is wrong.
- (2) The doctor did not do the right thing.
- (3) Therefore, ethical egoism is not correct.

This argument avoids the question-begging criticism and remains a serious criticism of ethical egoism. However, it is still not possible to analytically engage with the argument without further defining what is meant by ethical egoism. Without such a definition the truth of premise 1 is unknown; it must first be determined exactly what egoism would require the doctor to do before it can be determined whether by ethical egoism the doctor did the right thing.

§7.2 Defining Terms

Rachels, perhaps assuming egoism to be inherently flawed, does not define the doctrine he is dismissing. The nearest he comes to a definition is in a later article where, in response to criticisms made by T. Machan (1978), he writes "If ethical egoism is correct, then the right thing for anyone to do, on any occasion, is whatever would best promote his own interests." (Rachels, 1978, p. 426). Such a simple (partial) definition is insufficient and unlikely to be accepted by anyone claiming to be an ethical egoist or seriously defending ethical egoism as a moral doctrine. If Rachels' only claim is that selfish pursuit of one's own interests is a flawed doctrine then there would be little point in quarreling. However, it appears Rachels wants to discredit consequentialist egoism in whichever way it might be defined and establish that any such theory built on self-interest is pernicious. This

being the case, Rachels is also directly attacking the egoist theory I have been developing throughout Part 1 of this thesis, that:

- (i) (*Universal principle of ethical egoism*) If Φ is recognised as good in itself relative to Agent_n's worldview and acting to bring about Φ is the act that would maximise Agent_n's interests then Φ provides Agent_n with strong moral reason to, and justification for acting.
- (ii) (*Individual motivation for acting*) If Agent_n believes that ethical egoism provides preferable reasons for acting then Agent_n will, in most cases, be motivated to act in accordance with (i).

In order to rise to Rachels' challenge I will spend a little time further developing the egoist doctrine. I will begin with the principles above and develop them further by consideration of what Rachels might mean firstly, by *interests*, and secondly by what it means to *best promote* these interests.

§7.2.1 The Egoist's 'Interests'

Interests must be more than just what an agent desires or believes to be in the agent's best interests. To claim that an act would lead to the maximization of an agent's interests is to claim that an act, all things considered, would most probably lead to the maximization of the agent's interests. Undoubtedly in many cases it would be difficult for a third party to establish what was in an agent's best interests, even if that third party had some understanding of the agent's worldview. However, all I am claiming is that some acts are quite clearly not in an agent's best interests no matter what that agent might claim to the contrary. For example, it is quite clear that while an agent may claim it is in his interests to jump off a cliff (believing he can fly), or to drink poison (believing he is invulnerable to the effects of that poison), it is not in the agent's actual best interests to do so. The agent's belief in such cases is erroneous and acting in the manner the agent believes will lead to the maximization of the agent's interests is not the act that will actually accomplish this goal.

Further, it is not only impossible actions that are ruled out, for, although I may have an interest to be financially well off it does not follow that I have an interest in emptying my bank account and spending all the money on Lotto tickets. In fact, given the high probability that I would lose most of the money, it is in my rational

interests to do the exact opposite of what I believe it would be in my interests to do. Thus the egoist must not only balance long- and short-term interests, but these interests must also be pursued rationally. 'Rational interests' is a concept that is unlikely to be entirely sufficient for the task at hand but, for the sake of brevity, I will simply accept that we can identify a commonsense notion of rational interests where, for example, self-destruction, self-deception or acting in a manner that in all probability will not bring about the desired outcome is not rational.

Probability of success is also problematic. An agent may have a rational interest in accepting some risk if the actual payout, if successful, is suitably compensatory, and the agent does not stand too great a loss if success is not forthcoming. An egoist making little headway in saving to buy a house may argue that his interests lie in using the rent money to play roulette and win enough to buy a house, and this may well be true, but given the improbability of this action being successful the egoist's better interests appear to be paying the rent and not getting evicted. The loss (eviction) if the egoist is unsuccessful is too great a risk and the act (gambling) is clearly not in the egoist's rational best interests.

Interests then, are not the agent's immediate desires nor are they merely what the agent believes will lead to the maximization of interests. To claim that an act is in my interest, such that it will result in the maximization of overall interests is also to claim that the act I believe to be in my interest has a reasonable probability of success, while factoring in the negative consequences such that should the act be unsuccessful the negative consequences will be minimised.

§7.2.2 'To Best Promote'

Rachels' term 'To best promote' interests is unlikely to mean 'to advance a single interest' nor is it exactly synonymous with 'to maximise interests'. Most, if not all, individuals have a range of rational long- and short-term interests. An agent attempting to promote these interests will often have a choice of several actions all of which have a reasonable probability of success. However, promoting one or a specific range of interests will in many cases hinder others. Considering a specific, if extremely simplified, example:

Max has 12 different fields of interest and at this point in time two specific courses of action that he could follow.

Action #1 would result in ten of these interests being slightly furthered and two of them being frustrated (to an equal degree). Let's say the value of each action furthered is 10 and in that case Action #1 can be assigned a value of 80: Ten interests furthered (100) less two interests frustrated (-20).

Action #2 would further a single interest a great deal (let's say a value of 80) and frustrate none.

I take 'to best promote' to mean something like 'given all of an agent's interests, a specific act would promote rather than detract from the egoist's balanced (long- vs short-term) rational interests. That is, overall more of the egoist's real interests will have been promoted than have been demoted; in short, the egoist has maximised his interests. In the above example both actions produce a utility of 80, however action #1, while increasing more of the egoist's interests, actually detracts from two of them whereas action #2, while only promoting a single interest, detracts from none. Even though both actions produce the same overall utility I suggest an egoist ought to adopt action #2. Although the utility is theoretically equal I contend that the cost of sacrificing something the egoist has an interest in is greater than the minor gains made in other areas. I will argue this point more fully in Chapter 9.

Adopting Rachels' concepts the universal principle of egoism can be extended to:

(Universal principle of ethical egoism) If Φ is recognised as good in itself relative to Agent_n's worldview and acting to bring about Φ is the act that would maximise and would promote rather than detract from Agent_n's 'balanced (long vs short term) rational interests and that acting to bring about Φ has a reasonable probability of success (i.e. of actually bringing about Φ) and the risk, should success not be forthcoming, is minimised' then Φ provides Agent_n with strong moral reason to, and justification for acting.

§7.2.3 The Extended Argument:

Incorporating both the definition of egoism and support for premise 2, by way of an appeal to the best explanation, the argument under consideration has become:

(1.i) If ethical egoism is correct then the right thing for anyone to do, on any occasion, is whatever action would maximise and would promote rather than detract from an Agent's balanced (long vs short term) rational

interests, assuming such action has a reasonable probability of success and the risk, should success not be forthcoming, is minimised.

(1.ii) Within the context of Rachels' scenario, the conditions of (1.i) would be met by taking the woman's money.

(1.iii) The doctor took the woman's money.

(1) If ethical egoism is correct then the doctor did the right thing.

(2.i) Most people would have an adverse reaction upon witnessing a scenario whereby a doctor took unnecessary advantage of a helpless woman by taking her last twelve dollars.

(2.ii) The best explanation for why the majority of people had an adverse reaction to the doctor's act, given no other overriding explanations, is that the act is detrimental to the woman's (and perhaps the doctor's) flourishing and that most people recognise the wrongness (in this case) in acting to the detriment of the woman's flourishing.

(2.iii) [by 2.i/2.ii] that it is a moral fact within Reality_s that taking unnecessary advantage of a helpless woman is wrong.

(2) The doctor did not do the right thing.

(3) Therefore, ethical egoism is not correct.

Having fully extended the argument it can now be considered whether Rachels' conclusion that egoism is a pernicious (incorrect) doctrine necessarily follows.

§7.3 Machan's Objection

Allowing the second premise (the doctor did not do the right thing) to stand, Machan (1978) outlines a possible attack on the first premise (1) and more specifically the claim of sub-premise (1.ii) that the doctor's act would be the act promoted by the application of the principle of egoism.

Machan (1978, p. 422) claims:

What Rachels does not prove is that, under the circumstances, making himself twelve dollars richer, versus some other action the doctor might have taken, was to the doctor's advantage.

Machan argues that taking the twelve dollars is likely to have prevented the doctor earning more money, damaged his reputation, damaged the doctor's own good

feelings, and that these negative factors far outweigh the small monetary gain. Machan (1978, p. 423) goes on to claim that these factors bring (1.ii) into serious doubt and that it most likely would *not* have maximised the doctor's *balanced (long vs short term) rational* interests to take the twelve dollars. Rachels (1978) finds no problem in countering Machan's criticism pointing out that, *as a matter of fact*, taking the twelve dollars had no actual negative effects on the doctor's interests. His reputation did not suffer and the doctor would not have had good feelings from not taking the money (in fact he would never have given the woman another thought).

We have only the limited context of the case study and the facts as presented by Rachels to go by, but I find it difficult to accept Rachels' claim that taking the woman's last twelve dollars did not negatively affect the doctor's best interest. If the doctor's act was clearly recognizable as reprehensible, as Rachels claims it to be, and the story is widely known then we must assume that the majority of people in the town, affluent and poor alike, unless they are all of like-mind to the doctor, considered that the doctor had acted reprehensibly. However, if Rachels is correct then these same people, despite their feeling about the doctor's actions, continued to treat him in a manner that had no negative effect and caused no damage to *any* of the doctor's interests sufficient to offset an extremely small financial gain.

However, even assuming Rachels is correct and the doctor suffered no negative effects from taking advantage of the woman's predicament, Rachels has ignored Machan's central criticism. Machan is claiming that the act of taking the twelve dollars is unlikely to have been *the* act, given all the possible options open to the doctor, that would have maximised the doctor's overall (balanced long vs short term rational) interests. Rachels' analysis rests on a false dilemma: limiting the consideration of options to only taking or not taking the twelve dollars. There were numerous options open to the doctor (some of these are discussed in §7.6) and if we are to assess whether an egoist would approve of the doctor's actions and judge the actions morally right we need to determine whether taking the twelve dollars, rather than any other action open to the doctor, would have maximised the doctor's interests.

It is logically impossible to assess whether the doctor did the right thing (according to egoism, or any consequentialist theory) to the extent that we can clearly and uncontroversially show that action 'x' is the one that maximised interests. While we have reasonable information about how things have actually turned out for the doctor (that is, the doctor is twelve dollars better off and has suffered no noticeable negative effects) we can only make assumptions as to how things might have turned out had the doctor acted differently. To compare the set of actual consequences with the sets of possible consequences from different actions is pointless. The case can equally be made that if the doctor waived the twelve-dollar fee the result would be:

- (i) He was twelve dollars worse off and the action had no other effect on his interests.
- (ii) He was twelve dollars worse off and plagued by patients wanting free services.
- (iii) He was twelve dollars worse off initially but his reputation as a caring doctor eventually resulted in increased income.
- (iv) He was twelve dollars worse off initially but the woman felt so indebted that she carried out actions that greatly benefited the doctor.

I will return to an attempt at calculating consequences in §7.6. All I am claiming here is that (i) through (iv) are all logically possible outcomes of just one of the many other options open to the doctor. Given that it is extremely unlikely that agreement could be reached as to the consequences of all of the other actions open to the doctor we cannot even begin to calculate the degree to which they would help or hinder the doctor's overall interests. Further, if we cannot calculate the degree to which the other possible actions would have affected the doctor's overall interests we cannot know whether the action the doctor did take is the one advised by the egoist doctrine. The premise that "On that occasion, it would maximise the agent's overall balanced (long vs short term) rational interests to take the woman's money" is far from uncontroversial. Whether an egoist would assess that the doctor acted correctly in taking the woman's last twelve dollars depends on the values assigned and the outcomes projected. Machan's criticism, namely that Rachels has not 'proved' that the doctor's act is the one that would have maximised his overall interests, while not disproving Rachels' argument, calls it into serious doubt.

§7.4 Hypothetical Cases

Rachels does take on board something of Machan's criticisms and counters with the argument that it really does not matter if he is wrong about the specific details of this case. Hypothetical cases are as valid as real cases in examining moral theories and a hypothetical case can be created similar to the one presented, but where as a matter of stipulation taking the twelve dollars does result in the maximization of the doctor's overall interests. Thus, in this hypothetical case the doctor does the right thing according to ethical egoism, acts in a manner that we know to be wrong, and egoism is still shown to be incorrect. Premise (1.ii) is modified to:

(1.ii) Within the context of *some hypothetical scenario*, the conditions of (1.i) would be met by taking the woman's money.

However, if we accept Rachels' claim with regard to the force of the hypothetical scenario then it must be assumed that we can create further hypotheticals to test Rachels' thesis. Consider the following hypothetical scenarios:

t_1 : The doctor/woman scenario as at the time Rachels describes it.

t_2 : Exactly one year after Rachels' scenario it turns out that, as a direct result of having her last twelve dollars taken by an unscrupulous doctor, the woman forms a civil rights movement which results in higher wages, better health care and a general improvement in the conditions of everyone in the black community without any adverse effects on the white community generally and without any effect on the doctor specifically.

t_3 : Exactly one year after t_2 the woman is now held in extremely high regard and charges are brought against the doctor resulting in his disbarment from the medical practice.

§7.4.1 Assessing Hypotheticals – Egoism

At t_1 the doctor's overall rational interests are maximised, as a matter of stipulation, and so we can conclude that at t_1 , according to ethical egoism, the doctor did the right thing to take the money. At t_2 little has changed that affects the doctor's interests. Others are better off but unless the doctor has an interest in preventing others from doing well (and it is hard to imagine such an interest qualifying as rational) then taking the twelve dollars would still be assessed as a right action according to egoism. However, things have changed considerably at t_3 . The act of taking the twelve dollars, assessed as correct at both t_1 and t_2 , has now quite clearly turned out to be detrimental to the doctor's overall interests (the

loss of income alone from being disbarred would offset the initial gain) and by the doctrine of ethical egoism the doctor did *not* do the right thing in taking the twelve dollars.

If egoism is to be judged by examining if the doctor's act actually maximised his interests then the assessment of the doctor's act may, and in our hypothetical cases does, vary depending on the point in time at which the act is assessed. How then are we to assess objectively whether the doctor's act was right from an egoist perspective?

§7.4.1.1 Option 1 : Denying Temporal Importance

One logical, if slightly bizarre option is to deny that the temporal location of the assessor is of any importance. However, if this position is adopted and t_x is irrelevant the result is the somewhat absurd position that t_1 and t_3 are equally valid assessments. This being the case, according to ethical egoism:

At t_1 the doctor did the right thing.

At t_3 the doctor did not do the right thing.

And, if we deny any relevance to t_x

The doctor *both* did and did not do the right thing in taking the twelve dollars

This inconsistency makes it impossible for an egoist to assess the correctness or otherwise of the doctor's act or the correctness of any action under any consequentialist theory. Given that this position results in both a logical contradiction and in the inability to make any definite consequentialist judgments I am assuming, without further argument, that the temporal location of the assessor *is* important.

§7.4.1.2 Option 2 : Accepting All Temporal Assessments as Important

Another option would be to claim that assessment of the doctor's act from any and all temporal locations is relevant. If *at any point* a pernicious act is shown to be right according to the egoist doctrine then the theory is incorrect. This is a somewhat odd position for it allows that even if the egoist knows what all the consequences of his act will be (t_1 through t_3), and knowing this is compelled not

to take the money (his long term t_3 balanced interests are maximised), egoism (and indeed any consequentialist theory) is still proved to be incorrect judged solely on what would have maximised interests at t_1 .

1. As an ethical egoist, the act that maximises my rational balanced interests (t_3) is not taking the money.
2. As an egoist I don't take the money.
3. If assessed at t_1 or t_2 , by calculating the actual consequences to date, I ought to have taken the money.
4. Because [by 3] I ought to have taken the money at t_1 or t_2 , then even though [by 1] I don't actually take the money, egoism is incorrect.

This position results in the absurd conclusion that a moral theory is proven incorrect by assessment of an act no-one correctly applying the theory would take.

§7.4.1.3 Option 3 : Defining the Moment of Importance

The third option is to provide argument that some specific time is the position from which an accurate assessment can be made. In the given hypotheticals it seems t_3 is a far better position from which to judge the act than t_1 . If the promotion of balanced rational self-interests is the goal of the egoist then quite clearly by t_3 the doctor did not do the right thing. If temporal location is important then it is absurd to suggest that the doctor's act be assessed at t_1 when his real balanced rational interests lie at t_3 . However, for any hypothetical t_x another hypothetical t_{x+1} can be created that further modifies the outcome and is a truer indication of the doctor's real rational interests.

One method of locating an objective temporal location would be to judge the real scenario at the point of the doctor's death, at which point all of the actual consequences of the doctor's act (that have had any affect on the doctor) can be calculated⁸⁹. But, apart from the absurdity of only ever being able to determine after the agent had died, if any of the acts undertaken by an agent throughout that agent's entire life were right, this option leads back to the problems discussed at §7.3. It is unlikely any consensus could be reached as to what the consequences would have been had the doctor acted differently than he actually did. However, there is one other point at which the doctor's actions could be assessed, namely; at the point of the moral decision. At this point all of the probable outcomes are

⁸⁹ Possibly, though more controversially, the doctor's interests might continue past his death. However, I will not pursue this extra level of complexity.

equally hypothetical, and the possible actions can be assessed in terms of the likely outcomes. I will discuss this option further in §7.6.

Despite the importance of the problems raised by the temporal location of an assessor, Rachels can sidestep this issue by reformulating his hypothetical scenario, by adding the stipulation that, as a matter of fact, taking the twelve dollars does result in the maximization of the doctor's overall interests at all possible assessment points. At any t_x it is true that the act of taking the woman's last twelve dollars maximised the doctor's overall interests.

§7.4.2 Assessing Hypotheticals – Generalization

In adding further stipulations Rachels' hypothetical scenario becomes more and more removed from reality. It is hard to imagine any moral theory that could stand up to increasingly bizarre and ever more stringent hypotheticals. However, it is not my intention to mount an attack against the validity of using hypothetical scenarios to judge moral theories. Rather, I will simply show that Rachels' argument works too well. That is to say, it dismisses far more than just ethical egoism as wrong and leads to conclusions that even Rachels is unlikely to be satisfied with. The central premise of Rachels' argument can be generalised to:

Any moral theory (X) that compels us to commit an act (Y) where (Y) is a wrongful act determines that the moral theory (X) is itself objectively wrong.

If Rachels' argument 'If egoism compels us to *take the woman's last twelve dollars* where we know *taking the woman's last twelve dollars* is a wrongful act then this determines that egoism is itself wrong' is sound, then any argument of the same logical form is equally sound. Applying the generalization we can formulate the following argument (of the same logical form):

If utilitarianism compels us to *take the woman's last twelve dollars* where we know *taking the woman's last twelve dollars* is a wrongful act then this determines that utilitarianism is itself wrong.

Applying this argument to the hypothetical scenario t_2 (see §7.4) whereby the consequence of the doctor taking the woman's last twelve dollars resulted in a maximization of the interests of the majority proves, according to Rachels' argument, that utilitarianism is not a correct moral theory.

(1.i) If utilitarianism is correct then the right thing for anyone to do, on any occasion, is whatever action would maximise the ‘balanced (long vs short term) rational interests, that have a reasonable probability of success, of the majority of the agents involved.

(1.ii) Within the context of Rachels’ scenario, the conditions of (1.i) would be met by taking the woman’s money.

(1.iii) The doctor took the woman’s money.

(1) If utilitarianism is correct then the doctor did the right thing.

(2) The doctor did not do the right thing.

(3) Therefore, utilitarianism is not correct (is a pernicious doctrine).

No doubt the scenario is somewhat implausible and it is unlikely that such an event would actually occur. However, the scenario is logically possible and given Rachels’ insistence that hypotheticals are a legitimate test of moral theory then the presented argument is, by Rachels’ criterion, an acceptable test of utilitarianism.

The doctor’s act of taking the woman’s last twelve dollars is unchanged in the modified hypothetical and, according to both Rachels and our criterion of best explanation, is morally wrong. However, at t_2 the utilitarian would judge the act to be morally right, in that the interests of the majority affected by the act *have* been maximised. If Rachels’ argument is sound then, as utilitarianism recommends an act clearly recognizable as wrong, utilitarianism must also be a morally pernicious doctrine. Rachels’ argument equally dismisses egoism and utilitarianism as wicked and incorrect moral theories.

Such a dismissal of utilitarianism is likely to face far more rigorous objection than when the exact same criticism has egoism as its target, even though the argument is of the exact same logical form. Objections aside, perhaps it is the case that all consequentialist theories are morally pernicious? Rachels does not seem prepared to either advance this generalised form of his argument nor accept its logical conclusions and elsewhere (1993, p. 116) takes a much softer line against utilitarianism claiming not that utilitarianism is incorrect but rather that utilitarianism is not “a fully adequate theory”. Exposing the argument’s

generalised form does little except to show that in this context utilitarian and egoist theories stand or fall together. Either both are ‘incorrect’ or both are less than ‘fully adequate’. However, that utilitarianism and egoism stand or fall together is not a defense of egoism; it may be the case that both are equally flawed. I maintain that the problem is not in the theories themselves but in attempting to assess acts by reference to actual outcomes.

§7.5 Temporality Again – Actual and Probable Consequences

As discussed (§7.4.1) there is a general problem in determining the correct act by reference to the actual consequences at some specific point t_x . The problem is that the result of the assessment may vary depending on the temporal locations chosen. If we assume that ‘taking unnecessary advantage of a helpless woman’ is objectively wrong, or at least a moral fact within Reality_s then in assessing consequentialist theories against the hypothetical scenarios at t_1, t_2, t_3 (§7.4) we find that:

According to utilitarianism, taking the woman’s money is assessed right at t_2 and t_3 , but wrong at t_1 , and that:

According to egoism, taking the woman’s money is assessed right at t_1 and t_2 , but not at t_3 .

Both egoists and utilitarians will assess the act as right at t_2 but will disagree over the moral status of the act at t_1 and t_3 . Such temporal problems can be overcome if the actions required by consequentialist theories are judged at t_0 ; the point at which a moral decision is to be made. By judging consequentialist theories by the probable consequences of potential acts we consistently compare the full set of probable consequences, (the most probable outcomes of the range of actions open to the doctor) rather than attempting to compare a single actual consequence, at some arbitrarily chosen future point, against what might have been the case had other options been chosen⁹⁰. If right action is assessed on probable outcomes we can ask:

Given the information available to the doctor at t_0 , which action should an egoist assess as the morally preferable choice?

⁹⁰At this point egoism has evolved considerably from the doctrine Rachels’ initially criticises. However, as previously claimed Rachels appears to be saying that the entire set of egoist theories ought be classified as pernicious.

Perhaps this is why Machan (1978) phrases the question not as ‘Would an egoist judge the doctor’s act as right?’, but rather ‘Was Rachels’ Doctor Practicing Egoism?’ or at least ‘Was Rachels’ Doctor Practicing a coherent and consistent form of rational Egoism?’ Further if we are examining the action Rachels’ doctor took we ought not be considering:

- (i) With reference to the actual consequences, at some future point t_x , would the doctor’s past act be judged correct by an egoist?, and instead ought ask:
- (ii) With reference to the range of probable consequences that would ensue from the range of potential actions, would Rachels’ doctor’s decision to pursue the specific act of taking the twelve dollars be judged correct by an egoist?

As has been argued throughout this chapter, judging egoism by (i) is plagued with difficulties – the correctness of egoism varies depending on the temporal location of the agent. Further, the correctness of an egoist’s actions turns out to be little more than moral luck – assessment of an egoist’s act depends entirely on the fall of the cards and how events over which the agent has no control, actually unfold. However, (ii) provides a consistent means by which to evaluate consequentialist theories, while facing some difficulties in correctly evaluating probable consequences.

§7.6 Calculating Probable Consequences

Attempting to assign numerical values to the doctor’s interests and then to calculate the action that ought to be taken based on probable consequences is difficult and almost certain to be controversial. However, I think enough can be done to show that it is highly unlikely that any rational egoist would, after considering all of the possible consequences of the various options, choose to take the twelve dollars over some other action in the scenario Rachels describes.

At t_0 the doctor has just finished his very brief examination of the woman, has done nothing for her, has asked for his fee and discovered that the woman has only twelve dollars to her name.

At this point the doctor has numerous options: charge anything from one to twelve dollars, take the twelve dollars and inform the woman she will have to pay the balance with interest (and given Rachels’ description of the doctor we might wonder why he didn’t pursue this harsher option), as well as more generous

options such as waiving the fee or feeding the woman. However, there seems to be no point in obfuscating the calculations so let us assume the choice is between (a) taking the twelve dollars, and (b) waiving the fee. These are the same options Rachels falsely limits the consideration to in the real case study. However, in our analysis of the probable consequences of these two choices we have no knowledge of what the actual outcomes will be or what will, as a matter of fact, maximise interests. The doctor, if he is an egoist, *must* weigh up the probable negative and positive effects of these two acts and rationally determine which ‘would most probably maximise his overall balanced (long vs short term) rational interests’.

§7.6.1 Option 1: Take the money and run.

If the doctor decides to take the twelve dollars then the positive effect is instantaneous and easy to measure; the doctor is, as a matter of fact, twelve dollars richer. Assuming that the accumulation of wealth is a real and rational interest of the doctor then the gaining of twelve dollars will count toward maximizing his interests. There may be a few other positive effects; others trying to get a service for free will not approach the doctor, he may even gain some form of satisfaction from the woman’s suffering; but the primary gain appears to be the twelve dollars.

To act *as* an egoist the doctor must also weigh up the probable negative consequences of the act. Assuming that Rachels is correct and the doctor is the type of character who will suffer no pangs of conscience if he takes the twelve dollars, nor even give the woman another thought, then any negative consequence must be external to the doctor himself.

Machan (1978) suggests many probable negative consequences: It seems probable that the woman will tell others about the way she was treated by the doctor (and in the real case we know this happened as the story is widely known) and given that the doctor’s act is, according to Rachels, so clearly recognizable as reprehensible it seems equally probable that this will cause the doctor to lose some future business. Even if the doctor is the only doctor in town people are unlikely to visit his surgery unless it is vital to do so - it is doubtful someone with a flu would visit a doctor they believed might charge without adequately performing a service, or

indeed performing any service. While the impact might be small, if the doctor loses only one full paying customer (i.e. twenty-five dollars) then by taking the woman's twelve dollars he may have acted against his own interests; he is ultimately thirteen dollars worse off. However, showing that the consequence of an act has a negative total value (-13) is not enough. It may well be that all of the other options open to the doctor result in consequences with a lower total value and diminish the doctor's interests to a greater extent. It must be shown, if the egoist is to claim it is wrong to take the twelve dollars, that some other act will have a better outcome (that is a greater calculated value and a greater probability of maximizing the doctor's overall interests).

§7.6.2 Option 2: Egoistic Benevolence.

The second option open to the doctor is to waive the fees (remembering we are artificially limiting the scenario to taking, or not taking the money). Choosing this action has an immediate negative effect for the doctor; he misses out on twelve dollars. He may also feel that he has wasted 10 minutes of his time (the doctor's loss is actually -25, the full fee he would have received if he had seen a full paying customer). The doctor might also fear that if the story gets around he may face numerous others expecting free consultations to the detriment of his business. The doctor, being the sort of character he is, presumably gains no pleasure from helping the woman so any positive consequences must, once again, be external to the doctor.

The story that the doctor waived the fee of someone in need will, in all probability, have some positive effects. We can assume that the story will enhance the doctor's reputation (as a caring physician), regardless of whether the reputation is actually deserved or not. Even if the doctor cares nothing for his reputation he ought, rationally, to recognise that in most cases a good reputation is directly translatable into increased earnings. Further, he can limit the negative effects by making clear to the woman that this was an extreme case or by limiting his clients from poorer communities.

This leaves the question of how we assess these two options, from an egoist perspective, such that one can be labeled correct.

§7.6.3 Assessing the Options

The doctor's decision is not as simple as a choice between taking or not taking twelve dollars. Rather, the decision is between (a) taking twelve dollars and the possible (and I suggest probable) loss of future income from a bad reputation versus (b) waiving the fee and potentially increasing future gains (via increased custom) due to an enhanced reputation. If it is accepted that taking the twelve dollars (a minimal financial gain) runs the risk of decreased income and taking a small loss by waiving the fee (with the potential for future gain) then the rational decision, when long and short term interests are considered is not to take the money. It begins to look very doubtful that Rachels is correct in claiming "if egoism is correct then the doctor did the right thing". However, the calculations are controversial and more work needs to be done to justify the opposite claim that "if egoism is correct the doctor did the wrong thing" (Machan, 1978, p. 413).

While it seems more likely that the doctor was acting out of greed than that his actions were guided by any ethical theory, and doubtful that an egoist would judge the doctor's actions as correct it is to no avail to push this point home. Rachels' argument is still far from defeated and can be reformulated one last time into its strongest form.

§7.7 Rachels' Argument – The Final Formulation

Given that we have accepted the use of hypothetical scenarios, Rachels can reformulate premise (1.ii) so that a scenario is created that, when all of the probable consequences of all of the possible actions are considered, it will still turn out that the action that maximises the doctor's balanced rational interests is to take the money.

(1.i) If ethical egoism is correct then the right thing for anyone to do, on any occasion, is whatever action would maximise and would promote rather than detract from an agent's 'balanced (long vs short term) rational interests which have a reasonable probability of success and the risk, should success not be forthcoming, is minimised such that negative consequences would not unduly impact on the agent'.

(1.ii) Within *some hypothetical scenario and on consideration of the probable consequences of the possible actions it would most likely* be that the conditions set by (1.i) would be met by taking the woman's money.

(1.iii) In such a *hypothetical scenario* the doctor will take the money.

(1) If ethical egoism is correct then the doctor will do the right thing by taking the money.

(2.i) The majority of people would have an adverse reaction upon witnessing a scenario where a doctor took unnecessary advantage of a helpless woman by taking her last twelve dollars.

(2.ii) The best explanation for why the majority of people had an adverse reaction to the doctor's act, given no other overriding explanation, is that the act is detrimental to the woman's (and perhaps the doctor's) flourishing and that most people recognise the wrongness (in this case) in acting to the detriment of the woman's flourishing.

(2.iii) [by 2.i/2.ii] it is a moral fact within Reality_s that taking unnecessary advantage of a helpless woman is wrong.

(2) The doctor will not have done the right thing in taking the woman's last twelve dollars.

(3) Therefore, ethical egoism is not correct.

All I am doing now is accepting that there is some hypothetical scenario where upon consideration of the probable consequences of the available action the doctor's interests will most likely be maximised by taking the twelve dollars. Further, this hypothetical need not be far removed from reality. Possible scenarios which might result in the doctor's interests being maximised are: Where the doctor is leaving the country the next day and there is no way that what the woman says in her community can have any possible effect on the doctor, where the doctor is old and his short-term interests have far greater weight than his ever shortening long term interests, or where the woman has no friends to confide in and will soon die.

The valid conclusion of Rachels' (generalised) argument is that, in some contexts, consequentialist ethics will compel an agent to commit acts that are, by way of a best explanation, clearly wrong. Further, while egoism may meet the minimal requirements for being a theory of action guidance it fails to meet the additional requirement necessary for it to be considered a *moral* theory of action guidance.

That consequentialist egoism *does* in some contexts compel an agent to carry out reprehensible acts leaves the egoist with two options.

1. Accept that egoism is not a moral theory and instead claim it is a preferable theory of action guidance.
or
2. Show where our view of Reality_s is mistaken.
If a more consistent, better explanation can be provided that has greater explanatory potency then it may well be that our intuition and beliefs (conservatism) are misplaced and that acts believed to be reprehensible are in fact the morally right acts in the given context. In short, claim that the theory is correct but our data mistaken.
or
3. Accept that ethical egoism is not a fully adequate moral theory.
That is accept that Rachels' criticisms do point to a flaw in consequential egoism (and possibly all consequentialist theories), a weakness that while insufficient to show the theory fatally flawed do highlight the fact that some modifications are required.

Option (1) defines consequentialist egoism as something other than a moral theory and is unlikely to be adopted by anyone seriously defending egoism. I am doubtful that those pursuing option (2) will have any great success. It would require a dramatic change in our view of the world to accept that unnecessary acts that cause great harm, even if they do bring about a desired consequence, are morally correct. However, the third option holds some promise and in Part 2 I will argue that ethical egoism can be further developed into a value-based theory that overcomes all of the problems that consequentialist-based egoism faces.

VIRTUE AND SELF-INTEREST

Part 2

Nonconsequentialist Egoism

Chapter 8

Flourishing

§8.0 Introduction

My aim in Part 1 of this thesis was to examine the idea that an ethical theory could be founded in self-interest by way of an analysis of consequentialist egoism and the major objections to, and criticisms of, this theory. In doing so it was determined that to overcome the claims that egoism was inconsistent and therefore not an acceptable theory of any kind, the following axioms needed to be adopted:

- (i) The Good is relative both to an agent and to that agent's specific worldview.
- and
- (ii) The Good, once recognised, provides an agent with strong moral reason and justification for acting in a manner so as to promote the Good.

However, while this line of defence showed that egoism was a consistent doctrine it was insufficient to substantiate the claim that egoism was a defensible *moral* doctrine. Adoption of the above axioms, applied within a consequentialist framework, left egoism open to the claim it was a pernicious doctrine, and therefore the antithesis of a moral doctrine. Rachels' criticism (discussed at length in Chapters 6 and 7) is that while Φ may be Good relative to a specific agent and to that specific agent's worldview, that acting to bring about Φ is clearly wrong and any doctrine that labels acting in such a manner as good is a pernicious doctrine. While Rachels' aim was only to discredit egoism his argument demonstrates a flaw in *all* consequentialist theories; for any consequence-based theory, hypothetical scenarios can be created whereby acting in the manner deemed correct by the consequentialist theory requires an agent to act in a manner clearly recognisable as pernicious, and therefore the theory is (a) seriously flawed, or (b) not a moral doctrine at all.

We could continue to defend consequentialist egoism at this point by:

Denying the claims I made in Chapters 6 and 7, namely that moral facts exist (in the form of best explanations), and responding that while some acts may appear pernicious this judgement is simply mistaken.

Such a claim would require showing, in scenarios such as the one presented by Rachels, that unnecessarily taking a helpless woman's last twelve dollars (even though that twelve dollars would only promote the agent's interest to a small degree and with the knowledge it would do a great deal of harm to the woman) was not, as a matter of fact, morally pernicious. Such an egoist must bite the bullet and claim that within a specific context such an act was the morally correct action, and while such a defence is logically possible, as I previously noted, I am doubtful that those pursuing this option will have any great success. It would require far too dramatic a change in our view of the world.

An alternative strategy at this point would be to accept that consequentialist egoism is seriously flawed. However, to defeat consequentialism and more specifically consequentialist egoism is not to defeat egoism per se. I contend that it is the consequentialist aspect of familiar versions of egoism that results in the promoting of pernicious acts and that egoism can be successfully defended as a nonconsequentialist doctrine.

I will begin to develop the concept of flourishing, more commonly associated with virtue ethics, and examine the thesis that what egoism requires is that an agent acts so as to promote that same agent's flourishing. More specifically I will examine the thesis that:

- (i) *(Universal principle of flourishing-based egoism)* If Φ is recognised as good in itself relative to Agent_n's worldview, that is, if acting to promote (or bring about) Φ would lead to the agent flourishing, then Φ provides Agent_n with strong moral reason to, and justification for promoting Φ .
- (ii) *(Individual motivation for acting)* If Agent_n believes that ethical egoism provides preferable reasons for acting then Agent_n will, in most cases, be motivated to act in accordance with (i).

In defending this thesis I will, firstly, analyse the concept of flourishing and question what it means to flourish both as an individual and as a being of a certain kind, who has the capacity to flourish, and when such flourishing is of moral importance. My primary approach to these questions will be an examination of Hursthouse's ethical naturalist thesis and the claim that flourishing differs and

increases in complexity across four classes of beings; plants, animals, social animals, and rational beings. In doing so I will argue that Hursthouse's bottom-up approach to defining flourishing is flawed and that a better model, one that avoids the problems inherent in the bottom-up approach, is a two-factor account that defines flourishing in terms of:

- (i) what it is for a creature of a certain type to flourish
and
- (ii) what it is for an individual, that is for those creatures that have a welfare of importance to them, to flourish.

In Chapter 9 I will examine the works of Ayn Rand, one of the leading proponents of egoism. The thesis developed by Rand is often treated as a variety of consequentialist egoism. However, it is far from obvious that this is the best description of Rand's theory. Many of the examples presented by Rand do not appear to be designed to promote the best consequences for an agent (notably where Rand insists an agent *must* risk their life), and are not specifically egoist in nature.

Rand was not an academic philosopher, in as much as she did not submit work to journals or engage in academic debate with regard to her work. Instead, for the most part, Rand preferred to present her theories in fictional literature. The nearest Rand came to presenting an outline of her theory in academic fashion is in a series of speeches later published in the anthology *The Virtue of Selfishness*. This work is somewhat confused, at times appears contradictory, and contains few specifics of the thesis Rand was extolling.

I will attempt to unravel the basic premises of Rand's central thesis by examining Nozick's (1971) reconstruction of her argument and Den Uyl and Rasmussen's (1978) criticisms of Nozick's reconstruction. I have selected Nozick's reconstruction as it might be reasonably concluded that, as a libertarian, he would be sympathetic to Rand's position. Yet Nozick's reconstruction is deeply flawed and demonstrates many of the mistaken ideas held with regard to Rand's thesis. Rasmussen and Den Uyl's criticism of Nozick is useful in highlighting these flaws and in formulating a more complete reconstruction of Rand's theory. My goal will not be to accurately reconstruct Rand's argument or to defend it. Rand's

principles do not, in themselves, demonstrate a complete theory, nor do I believe they are supposed to, as Rand was articulating principles to live by, not attempting to develop a fully structured system of ethical principles.

It is via Rand's basic principles, together with the axioms of egoism (Part 1), and the two-factor account of flourishing (Chapter 8), that I will develop a defensible theory of nonconsequentialist egoism. Hunt (1999) has already taken the first steps in this direction in his thesis of Flourishing Egoism which is itself built upon the foundations laid by Rand. It is to an examination of flourishing that I will now turn my attention.

§8.1 The Concept of Flourishing

A common dictionary definition of flourishing is: "To be vigorous or in good health; to thrive; to be at the height of success...." (Grollier Webster, 1974). Common definitions are seldom useful for philosophical purposes, and this definition is insufficient in determining answers to questions such as: 'What exactly does it mean for X to flourish?', 'Is X, at the height of success but in poor health, flourishing?', and 'Is an agent who suffers in one area but excels in others really flourishing?'.

However, putting such questions aside for the time being, the definition does seem to capture what we mean by flourishing in its simplest sense. Flourishing, if by flourishing we mean to thrive, to be a successful X, can be applied to anything that is animate, for example any living thing (cancer cells, mould, viruses, plants, animals, and humans), connected groups of living things (communities, ecosystems and species), and some non-living things (artificial intelligence or, perhaps, viri within a computer system).

Few people would object to the claim that a multitude of individuals, groups and systems can flourish. That is, X as a specific kind of thing can flourish or otherwise, as an entity of that type. Further, it appears equally clear that what we mean by saying that X is flourishing differs, depending on the complexity of the individual or system to which X refers. To say that a human is flourishing seems to imply more than simply that the person in question is healthy or has been

successful at some specific task. However, for animals, lower animals at least, being healthy and thriving, being a successful creature of that kind, may perfectly denote what it is for that organism to flourish. Similarly, the flourishing of an adult human is more complex and more difficult to describe than the flourishing of an infant or baby.

While possible, it seems unlikely that flourishing is of moral importance whenever and wherever it occurs. It appears equally unlikely that the axiom I introduced in Chapter 6, “If in a given context (Y) an act (X) would be detrimental to flourishing then we would expect, all things considered, that in most cases anyone directly observing X in context Y would experience an adverse reaction”, is sufficient. I still hold this axiom to be true within a limited scope. It is not the case that witnesses to numerous and everyday acts that *are* detrimental to flourishing suffer an adverse reaction. Pouring bleach (X) on bathroom mould in order to make a bathroom hygienic (Y), while quite clearly detrimental to the flourishing of the mould, is unlikely to produce an adverse reaction in all but a very small minority of persons witnessing such an act. Likewise, the act of eliminating a bacterial infection or pulling weeds from a garden, while detrimental to that organism is unlikely, in most cases, to be the cause of adverse reactions.

Nor does it appear the axiom can be saved by the adoption of an anthropocentric approach and speaking only of human flourishing; sport provides numerous examples where the witnesses of acts clearly detrimental to the contestants’ flourishing do not cause adverse reactions in an observer. Rather, it would seem the opposite is true and the witnesses take pleasure from the suffering of others⁹¹. Further, any attempt to limit the scope of the axiom to humans appears destined to failure because it seems *ad hoc*. One of the strengths of the axiom is exactly that it may include the actions of non-human creatures. In the case of Stingl’s apes (§6.9) the explanation that the female apes reacted upon witnessing an act (excessive punishment) by the dominant ape appears to be the best explanation

⁹¹ I am not referring to suffering in terms of the team suffering by losing but more like the excitement generated when cars crash during racing events, or a particularly vicious tackle during a rugby game. It seems that these pictures are always reported on the front of papers, get the most replays on TV, generate the most conversations and that people, in general, take some form of pleasure from viewing and talking about these events.

available, given the predictive leverage that can be gained from adopting the premise:

If, in a given context (apes' tribal hierarchy), an act (excessive punishment by the dominant ape) would be detrimental to flourishing then we would expect, all things considered, that in most cases any ape⁹² directly observing X in context Y would experience a negative reaction.

Perhaps clues to the question of the correct scope of the axiom can be uncovered in determining a more accurate definition of what it means to flourish and it is to this initial question that I will now turn my attention.

§8.1.1 Flourishing and the complexity of life

Hursthouse (1999) argues that although all living things can flourish, what constitutes an entity's flourishing increases as we climb the evolutionary ladder and as life-forms increase in complexity. Hursthouse begins this climb up the evolutionary ladder with plants. A plant is evaluated as being a good example of its kind if it meets two criteria, and evaluated as a good member of its species if it meets two ends. A plant flourishes if it meets the criteria of:

- (i) having good parts (root system, stalk, leaves etc), and
- (ii) having good operations (extracting nutrients from the ground, photosynthesis, etc.) (Hursthouse, 1999, p. 198).

Further, a plant can be said to flourish if it meets the two ends of:

- (a) individual survival, and
- (b) promoting the continuance of its species (Hursthouse, 1999, p. 198).

An apple tree with (i) a strong trunk and good root system that (ii) extracts the nutrients it needs from the soil, (a) survives a lifespan typical of trees of its kind and, (b) produces fruit at the right time of year that ripens and contains fertile seeds can be said to flourish in the manner that plants (and more specifically apple trees) flourish.

The flourishing of lower (non-social) animals (as individuals within the species) is evaluated by the same criteria as plants i.e., (i) and (ii), together with the additional criterion that they:

⁹² I have replaced 'any one' with 'any ape' here on the grounds it may not be possible for non-apes to fully recognise what is occurring or to understand the specific worldview of apes.

- (iii) Act well. That is they act well in a manner that is characteristic of the species.

In addition to (a) and (b), animals flourish if they act towards the following end:

- (c) The “characteristic freedom from pain and (where appropriate) characteristic pleasure or enjoyment” (Hursthouse, 1999, p.200).

A bear might be said to flourish if it (i) is fit, has strong teeth and claws such that (ii) it is capable of hunting and further (a) the bear is successful at hunting and feeding, (b) it breeds and produces young for the continuance of species, and if (iii) the female bear provides food for and protects her young, as is characteristic of that species, while (c) being free from pain that is uncharacteristic of a bear (such as some form of physical deformity). The jump from plant to non-social animals is large and, except for being non-social, it would seem that bears have more in common with higher animals (perhaps all mammals) than other creatures. While the distinction between insects and mammals seems significant the line of demarcation between higher and lower animals appears somewhat arbitrary.

The flourishing of higher social animals is judged against the three previous criteria and ends, plus their desires and emotions, toward the “good functioning of its social group - in the ways characteristic of the species.” (Hursthouse, 1999, p. 202). The society of apes discussed by Stingl (§6.9) provides a perfect example; the apes in this social group monitored the actions of other apes within the group, and where necessary acted to modify that behavior towards the end of the good functioning of the ape social group such that the social group flourished.

The idea of introducing new criteria as the forms of life increase in complexity has merit but I am dubious that a naturalistic augmentative ‘bottom-up’ notion of flourishing is workable. While in many cases it is possible to judge flourishing by Hursthouse’s criteria, at other times it appears that the criteria are inconsistent or contradictory. Take, for example, the plight of the inhabitants of Philip Island: rabbits were introduced to Philip Island and with no natural predators they flourished.... and continued to flourish until the point of extinction. In this case to do what was characteristic of individual rabbits, to flourish (eat and breed), was the exact thing that prevented the continuance of the species (in this locality at

least). Perhaps some tale can be told to accommodate such occurrences, however it appears that in many cases the natural inclination toward individual survival (criterion a) directly conflicts with survival of the group (criterion b). I fail to see how any creature can be judged against an end, the attainment of which would mean acting against that creature's natural inclination.

Climbing further up the evolutionary ladder leads us to humans, beings that are "not only social but also rational" (Hursthouse, 1999, p.206). According to Hursthouse the difference between higher animals and humans beings is that humans (as a rule) act from reasons whereas all animals act only from inclination. That is, humans can (and, if it is to be considered a norm, mostly do) act from a base of rationality and reason a course of action in respect of the current situation rather than, as animals supposedly do, acting directly in response to the current situation. Humans have character traits and, according to Hursthouse, these are in some sense natural (in as much as they are determined via the naturalist process) and for a human to flourish requires not that they act in a specific manner (eat, breed, etc.) but that in whatever manner they act they act well, that is, they act in accordance with the virtues.

While I accept that humans can act well or otherwise I find the line of demarcation drawn between animals and humans to be both arbitrary and problematic. Having already related the story of Stigl's apes I find it difficult to see how such an explanation, purely in terms of inclination, can account for the consistency of the apes' acts, a consistency not just in how they act as individuals or within a social group, but how those acts are modified according to both context and the reactions of others. A far better explanation (with greater predictive leverage) is that these apes acted from rationality and that in this particular story the female apes' actions were reasoned. That is, the female apes acted rationally in assessing the justified level of harm (punishment) that can be handed out by the dominant ape in a specific context. If there is any difference between a dominant ape punishing others within the tribe (in what appears to be a justifiable context) and a human mother punishing her child to a justifiable extent (other than that we can understand the human mother when she later provides justification for her actions) or between the female apes assessing the punishment

dealt out by the dominant ape (and taking action if it is not) and one parent assessing if the punishment of a child by the other parent is justified (and taking action if it is not) then I am unable to determine what that difference is.

It seems to me that the claim that all animals act purely from inclination is wrong. Further, I am unsure of the extent to which the ability to provide reason after an event is any evidence that the initial action wasn't just from inclination. I am not claiming that animals act from reason, only that it is not clear all animals act only from inclination or that humans act from reason (as opposed to later rationalizing actions from inclination). The very concept of rational beings is vague and I am dubious it can work to separate humans from animals in the manner that Hursthouse believes, and results in the earlier criterion being conveniently set to one side when it comes to talking about humans.

The case study about the apes might be dismissed, or called into doubt, if it was an isolated one, but this is not the case. De Waal, reporting on a study of baboons, relates the story of a troop where the dominant males all died after eating meat from a dump (the surviving baboons had been too subordinate to attend the dump brawling) (Cited in Angier, 2004). What was unusual is that none of the surviving males assumed the dominant role, rather there was "a cultural swing toward pacifism". Many of the original survivors have since died yet the genial style has been maintained. As baboons from other troops try to immigrate and act in the usual dominant manner (to locate their hierarchical rating within the troop) it appears they are quickly instructed "we don't do things like that around here" and are initiated into the "unusual customs of the tribe". It would seem that the ethos of the troop is rationally cultivated and that the troop bands together to ensure that the peaceful lifestyle continues. Once again it seems part of a better explanation that rationality plays a role and that the troop has a rationally-based preference for a peaceful lifestyle and, due to the preference, teaches immigrants how to fit in with that lifestyle. Once again I fail to see any reasonable explanation to show these baboons are acting purely from inclination.

Perhaps it might be argued that these apes are simply higher up the evolutionary scale and are approaching rationality, but examples of animal actions that appear

to involve some degree of rationality are not limited to apes. Many animals thought to be basically unintelligent (lemurs in this case) have the capacity for rational intelligence. A recent research project showed that lemurs have the ability to remember complex sequences of patterns and to continue such sequences where other lemurs leave off (a task that is often claimed to show the ability for rational thought and intelligence). However, lemurs (much like humans) only work when there is something available that they want - “they’ll work for a couple of hundred trials because they want these sugar pellets, even though we don’t deprive them in any way” (Brannon, 2004)⁹³. It seems lemurs, previously considered to be unintelligent lower-animals, (i) make choices whether to work based on the degree to which they desire the offered reward, and (ii) carry out tasks that demonstrate both intelligence and rationality.

Such examples are not limited to quadrumanous primates. Vicki Hearne (1986) tells the true story of Fritz, a Doberman police dog:

Fritz was the partner of Philip Beem, who can’t be described as anything but a bad-assed cop. One night, Officer Beem stopped a young black woman for jaywalking and started clubbing her with his nightstick, for the sheer fun of it as near as anyone could make out. (There were witnesses.) Fritz attacked - not the woman, but his policeman partner, and took his club away from him emphatically.

Fritz was a well-trained police dog and consistently followed the commands of his partner yet in this circumstance he went against this training and instead protected the woman. There are many possible explanations for the dog’s act but these competing explanations are hard pressed to better the explanation that the dog made a rational choice. Fritz recognised something as wrong in Officer Beem’s action and took steps to prevent something clearly detrimental to the woman’s flourishing. A competing explanation would need to better explain why the dog protected the victim in this scenario but in every other aided the officer.

Likewise, journalism contains numerous stories, many with eyewitness confirmation, of animals that overrode their inclinations to run from fire and instead entered buildings to save people (sometimes complete strangers).

⁹³ Cited in *The New Zealand Herald* World News (Reuters, 13th May 2004, Greed Spur in Animal Intelligence).

Similarly, dolphins have been known on many occasions to put themselves at risk to save humans from sharks. My point here is not to deny that rationality adds an extra layer of complexity when describing what it means to flourish, only that the line Hursthouse attempts to draw between the flourishing of social animals and human flourishing is far less distinct than she appears to claim.

I am not claiming that animals act from reason as a norm. It is extremely doubtful that the Philip Island rabbits that drove themselves to extinction acted from, or were even capable of, rationality; but then, by the same token, it might be questioned whether humans who are increasingly laying waste to their environment, and act in a manner that restricts human flourishing, are acting rationally. I am claiming that rationality, or what appears to be rationality for all intents and purposes, exists on a continuum, a continuum that does not consist exclusively of humans. Drawing a division between human flourishing and non-human flourishing seems unsupportable and, I will argue, is an unnecessary distinction. The 'natural' elements of flourishing can be accounted for by a single criterion.

Returning briefly to Hursthouse's ethical naturalism and her notion of human flourishing, it seems that the augmentativeness of 'means and ends' suddenly ends when we come to talk of humans, in that rationality overrides all the previous criteria. For lower animals the criteria for flourishing are the same as for plants, but with the addition of an extra criterion. For social animals the criteria are the same as for lower animals with the addition of an extra criterion. However, when we progress to humans the entire set of earlier criteria give way and, somehow, character traits are of sole importance. This would be reasonable if such character traits incorporated the other criteria, however it seems that rather than incorporating the earlier criterion it allows that, through reason, they be overridden.

We are asked to accept that homosexuality (or celibacy) in animals points to a defect, no matter how the animal acts as a member of the pack (disregarding any character traits of the animal), as it goes against the end of the continuance of the species. However, we are then asked to accept that human homosexuals are not

defective human beings because what is at stake is not “sexual activity or orientation, but character traits” (Hursthouse, 1999, p. 214) regardless of the failure to continue the species. Unless a clear and relevant line of demarcation can be made between humans and (higher) animals such a position is inconsistent. Even if a clear division exists between humans and animals it remains unclear why the earlier criteria, attributed across plants, lower and higher animals, are completely overridden. According to Hursthouse rationality accounts for this difference in that “it is quite certain that it is primarily our acting from reason, well or ill, rather than those occasional actions we do ‘from inclination’, that makes us good or bad human beings in the ethical sense” (1999, p. 206). This may be, but if the ability to reason overrides those previously determined criteria of flourishing then it does so for any creature that has the ability to reason regardless of its rung on the evolutionary ladder or, if animals are subject to those criteria, regardless of any ability to reason, then so are humans.

The claim that humans act primarily from reason, rather than merely having the capacity to formulate reasons for having acted, seems insufficient. If indeed it is our acting from reason, well or ill, that determines if we are good or bad human beings then we need some means by which to evaluate those reasons by which we act. It cannot simply be that these reasons match with what a presupposed virtuous person would do, because the virtues are supposed to have been reached via the investigative process of ethical naturalism. That is, ethical naturalism is supposed to be something that provides the criteria for a particular character trait being a virtue.

If ethical naturalism is correct then it would seem that the best criterion for evaluating reasons would itself be based in ethical naturalism. That is, good reasons would in most cases support the previously determined criteria and ends, bad reasons would do otherwise. Thus any reasoning that went against the previously determined end of continuance of the species was not a particularly good reason. Human homosexuals may exhibit the virtues of being caring and temperate etc (assuming that ethical naturalism can support these virtues) but, under ethical naturalism, they must still be judged as defective human beings unless we accept as good reasons those that go against the previously established

criteria and ends and in doing so invalidate the entire augmentative project. The alternative, holding that human homosexuals are good people (assuming they hold the relevant virtues), is to deny that the end of continuance of the species is a means by which to judge flourishing at any level and again invalidates the augmentative project.

Rationality does seem to make a difference to what is denoted as flourishing for a particular individual but it is unclear who is included within the boundary of rational beings. A charitable interpretation of rationality will see a large number of animals included within its boundaries, whereas a restrictive interpretation will see many humans excluded (not just babies and the severely mentally impaired or comatose, but those with low IQ's). Definitions of rationality are certainly possible but it is unclear why any specific definition should be accepted over another that broadens or restricts the boundaries to a greater extent. The capacity to act from rational thought seems certain to impact on what it is for that individual to flourish. However, it seems dubious that these individuals are necessarily human or, that acting from rationality alone determines what it is to be a good or bad human being.

§8.2 Flourishing – A Two-Factor Account.

While I have argued that the naturalist augmentative system that Hursthouse employs is flawed, I have considerable sympathy with her final position and agree that what it is to flourish increases with complexity parallel to the increase in complexity of forms of life. However, I would claim that flourishing is not limited to living things, that flourishing need not be the same (or of the same type) for every member of a species, nor that the increase in complexity reaches its apex with humans. Ecosystems can flourish, and the complexity of that flourishing goes beyond what it is for any individual to flourish and is far more than the aggregate of individual flourishing. It is not the case that if every individual flourishes then the ecosystem flourishes.

Hursthouse's project is flawed in its 'bottom-up' approach. Progressing from the flourishing of plants, through animals, and finally to humans introduces inconsistencies and contradictions that are difficult, if not impossible, to

overcome. Further, when it comes to talking of good or bad in relation to humans, Hursthouse's theory relies on an undefined notion of rationality and requires the abandonment of the previously extolled criteria, used to judge whether plants and animals were good examples of their kind, and thus the abandonment of the augmentative process.

I agree with Hursthouse, and in doing so concur with Anscombe, Geach and Foot, that 'good' does not change its meaning when used in relation to species of plant, animal or human. However, this 'good' is neither augmentative nor based on a 'bottom-up' analysis of what it is for an evolutionary class of beings (namely; plants, lower animals, higher animals, humans) to flourish. While 'good' does not change its meaning, the answer to what is 'good' or more specifically what is it to flourish depends on the subject of the question. (a) What constitutes human flourishing (generally), (b) what constitutes human flourishing in some vastly overpopulated region of the planet, (c) what constitutes human flourishing in New Zealand, and (d) what constitutes my flourishing, will result in different answers.

I will demonstrate that a consistent notion of flourishing can be developed by adopting a two-factor account. The first factor is based on a holistic approach taking the ecosystem as a whole as the determining criterion for what it is for a creature to flourish, both generally and within a specific context. Such an approach may seem at odds with the apparently individualistic nature of egoism but, as I will later go on to demonstrate, this is not the case. The second factor is based on rationality; any individual that has a welfare that is of importance *to them* can flourish, in promoting that welfare, and can be harmed when that welfare is put at risk.

§8.2.1 The First Factor: Eco Flourishing

Aldo Leopold in his essay 'Ecocentrism: The Land Ethic' formulated the thesis that right action is directly related to, or at least derived from, what is good for the ecosystem, the biotic community as a whole. His controversial central claim is "a thing is right when it tends to preserve the integrity, stability, and beauty of the biotic community. It is wrong when it tends otherwise" (Leopold, 2001, p. 125).

This thesis has been widely attacked on the ground that it is ecofascist, that is, it requires acting for the good of the whole at the expense of the individual. Regan demonstrates the ecofascist nature of the theory using the following scenario:

If, to take an extreme, fanciful but, it is hoped, not unfair example, the situation we faced was either to kill a rare wildflower or a (plentiful) human being, and if the wildflower, as a 'team member', would contribute more to 'the integrity, stability, and beauty of the biotic community' than the human, then presumably we would not be doing wrong if we killed the human and saved the wildflower. (Regan, 1983, p. 362).

I will argue that it is possible to interpret Leopold in terms of flourishing and to adopt a naturalist approach that is top-down, focusing on the flourishing of the biotic community. Such an interpretation lessens the ecofascist nature of the thesis and at the same time overcomes the problems highlighted in Hursthouse's bottom-up approach to defining flourishing.

What is it for a certain form of life to flourish in a given context? A form of life can be said to flourish if, after examining the biotic community in which it exists, it is determined that, all things being equal, the form of life in question is acting in a manner that will lead to the integrity, stability and beauty of the biotic community.

Such a position avoids some of the pitfalls in Hursthouse's account. While the rabbits on Philip Island flourished in the simplest sense of acting in a manner characteristic of rabbits for the continuance of the species they were not good rabbits per se because they were out of balance with the ecosystem to an extent that would ultimately lead to their extinction. The rabbits, while breeding and multiplying, were not flourishing as rabbits (in the true sense of flourishing) because the actions of the rabbits were such that the only possible result was extinction and extinction is not a function of flourishing.

A holistic position also avoids the result of labelling an individual of a species, who acts outside of the norm for that species, as defective. A (human) homosexual is not defective if the actions of the persons involved are not detrimental to the biotic community in which they exist. By the same token animal homosexuals are not defective unless that homosexuality is in some way

detrimental to the biotic community. If a species (necessary to the flourishing of the biotic community) was low in number, be it human or not, such that the sexual orientation of individuals within that species put the species at further risk then such action would illustrate a defect (no matter what other characteristics they possessed). Where sexual orientation does not put the species at risk such action does not necessarily point to a defect. This holistic position has the benefit that (i) the exact same criterion is applied in all cases, and (ii) it avoids the inconsistency in Hursthouse's theory whereby animal homosexuality points to a defect (in failing to act for continuance of the species) but human homosexuality, despite failing the same criterion, does not.

Similarly, a 'free-rider' wolf is not necessarily defective. Hursthouse claims that "Wolves hunt in packs; a 'free-rider' wolf that doesn't join in the hunt fails to act well and is thereby defective" (1999, p. 201). However, if the pack has sufficient numbers to hunt and there is sufficient food to go around there is no reason why every wolf must hunt (in order to be considered a good wolf). Rather, it may be the case that, in order to maintain a natural balance within the ecosystem, should a pack grow too large, free-riders are not only nondefective but a necessary part of the pack. I do not know enough about the society of wolves to comment further. However, it does seem possible that the condemnation as a free-rider of a wolf that does not join a hunt has certain anthropocentric overtones; we tend to object to so-called free-riders in human society.

Humans (as a species) are good per se when they adapt and fit into the ecosystem; when rationality is used to carve a place in an ecosystem while minimising disruption to the environment. A good human, in this respect is no different from a good plant or a good animal with the exception of having a greater capacity to adapt and control through the use of rationality.

From the holistic perspective the question, "Whose flourishing matters?" does not arise. If flourishing matters it matters wherever it occurs and such flourishing is good, or individuals can be said to be flourishing, when their actions lead to a well functioning biotic community. In terms of the whole the flourishing of mould, plants, animals, and humans is all equally important. However, the question

“Whose flourishing matters?” does become relevant when an individual’s welfare is at stake, a welfare that is of importance to that specific individual.

The flourishing of an individual is not entirely captured by reference to ecosystems and a second factor is required in order to expound “What constitutes an individual’s flourishing?” The answer to this question can, quite obviously, be at odds with what is it to be a good human in this specific context.

§8.2.2 The Second Factor: ‘My’ Flourishing

Every living thing can flourish, and that flourishing can be determined on the grounds of contribution to the working of the biotic community. However, there is a distinction to be made on an individual level; that is, between those that can flourish and are in some way aware of and affected by the degree to which they flourish and those that can flourish but have no awareness of that flourishing. At a holistic level flourishing matters because if it should fail the ecosystem and everything in it fails as well. However, for those individuals who are aware of their own welfare and are, in some way, harmed when they are prevented from flourishing, that flourishing becomes of important moral concern.

The first question to be answered is ‘Who has a personal interest in their own flourishing?’ An initial answer might be humans, but such an answer will not stand up to close analysis. Not all humans have either an interest in, or awareness of, their own flourishing. The severely mentally incapacitated, comatose, young children and babies are obvious exceptions. Nor is the answer *human as a norm for that species* acceptable, either (i) some set of criteria separates humans from nonhumans, in which case the criteria, not *human as a norm for that species*, determines who has an interest in their own flourishing or, (ii) the answer is the speciesist naming of members of our own species as the only ones capable of having an interest in flourishing purely on the grounds of being of our own species. I will not attempt an analysis of who can flourish, as a related question has already been debated in great depth for numerous years under the guise of the extent of the moral community. Some philosophers, such as Kant (1949), have argued that the moral community consists only of rational beings, a position that results in the exclusion of some humans. Others, (notably Regan (2001) and

Singer (1976)), have attempted to extend the moral community beyond humans to include some animals. I follow Regan and claim that when we are talking about an individual's flourishing we are talking about an individual's welfare. Any individual that has a welfare that is of importance *to the subject*, in Regan's terms any "experiencing subjects of a life", can flourish and has an interest in that flourishing. Regan (2001) defines an experiencing subject of a life as a conscious creature having an individual welfare of importance to them whatever the usefulness to others. Any creature that wants and prefers things, believes and feels things, recalls and expects things has a welfare of importance and can flourish or otherwise.

For the purposes of this thesis it is not necessary to determine exactly who the *experiencing subject of a life* is, only to note that anyone who qualifies as such has a specific interest in their own flourishing. From an egoist perspective, though the egoist recognises others have an interest in their own welfare, it is the egoist's personal welfare that ultimately dictates action.

The two-factor account of flourishing I am proposing is:

1. (Environmental factor) For type X to flourish type X must be in balance with its place in the biotic community.

This notion of flourishing can be applied to species, communities, living and non-living things and is determined in the same manner on each and every occasion. A pack of wolves is good (is flourishing) if it has not grown too large (such that food is in short supply) and that there are sufficient wolves to hunt (not that every wolf hunts) so that pups are cared for. A biotic community flourishes when everything within that community is in harmony. The same applies to humans qua human beings, the harmony is disrupted where populations have increased beyond the carrying capacity of the land. To the extent that the mere existence of those humans is detrimental to the environment then those humans (as a species) are not flourishing, regardless of whether they act rationally or not.

2. (Individualist factor) Any X that has a welfare that is of importance to X can flourish when that welfare is cared for.

Any individual who has an interest in their own welfare can be said to flourish or otherwise when that welfare is cared for. This, to a limited extent, relies on rationality, although to what degree is debatable. All that is required is that an individual has an interest in their welfare and can choose to act, in a manner that takes that welfare into account.

Axioms (1) and (2) do not contradict each other, though they may produce inconsistent answers, as (1) judges whether the type X is flourishing, whereas (2) judges whether X is flourishing as an individual. We may judge, by (1), that humans in some famine-struck area of the world are not flourishing as humans qua human kind. However, that does not exclude the possibility (though it may place severe limitation upon it) that, by (2), the specific individual is flourishing in that despite the hardships, the individual manages to overcome adversities and promote their own welfare. At this stage (2) may sound like a return to a limited form of consequentialist or instrumentalist egoism, namely, that an agent ought to act so as to maximise that agent's welfare. This is not the case and, as I will demonstrate in the following chapters, acting for one's own welfare is not the result of a consequentialist calculation.

A lot more needs to be said of how this two-factor account of flourishing can be applied without leading to some absurd conclusions. In Chapter 10 I will argue that the environmental factor (1) necessitates that humans, if they are to have the best chance of flourishing, adopt certain virtues. However, I will delay such an analysis as I wish to first examine the ethics of Ayn Rand, through the eyes of Nozick (1971), Den Uyl and Rasmusen, (1978), Hunt (1999) and Rand herself, as I believe doing so will provide insights into what it is for an individual to have a personal welfare, a welfare that is steeped in the individual's personal values. In doing so it will become clear how acting for an individual's own welfare differs from the consequentialist's acting so as to maximise the agent's own welfare.

Chapter 9

Rand's Ethics

§9.0 Introduction

Rand is pretty much a pariah of the philosophical community, perhaps due to her move toward presenting philosophical doctrines through fiction, perhaps because she was, at times, openly hostile towards analytic philosophy, or perhaps simply because the positions she held (or that were attributed to her) are extremely unpopular. Certainly, her work does not meet the rigorous standards of analytical philosophy and at times contains much that seems inconsistent or contradictory; but this alone is not reason to shun her entire work. The exact same charges of inconsistency or contradiction might be laid against the work of many respected philosophers.

It may seem odd to introduce Rand's theories so late in a thesis on self-interest. However, the discussion to date has focused on consequentialist egoism and I do not believe Rand's work is best classified in this manner. The first indication that Rand does not extol consequentialist egoism can be found in *'The Fountainhead'*; Roark, an architect down on his luck and in dire need of work, has submitted plans for the Manhattan Bank Building, and is currently waiting on the decision of the board with regard to his designs. The board eventually agrees to contract Roark but on the condition that they will alter his designs. Despite being on the brink of destitution Roark refuses to accept such conditions. When quizzed on his decision by a board member, who asks, "Do you have to be quite so fanatical and selfless about it?", Roark responds: "That was the most selfish thing you've ever seen a man do". By the use of the term 'selfish' Roark indicates that he was acting from within an egoist framework, yet at face value his actions appear decidedly non-egoistic; Roark acts in a manner that is detrimental to his self-interest, in terms of consequences, in that he turns down a job that will provide him with the means he needs to survive. Roark's self-interest was based not in the consequences of actions but in the preservation of his personal values, values that included a love of architecture and values that he would have to compromise were he to allow his designs to be altered.

Further indications that Rand's ethical theory⁹⁴ is not consequentialist in nature are apparent in her essay on the '*Ethics of Emergencies*', in which Rand discusses how one ought to act when faced with a scenario that is not "metaphysically normal" (for a specific individual). Rand uses the term "metaphysically normal" to refer to any "unchosen, unexpected event, limited in time, that creates conditions under which human survival is impossible" (Rand, 1964b, p. 54), for example natural disasters, or being involved in a plane, ship or car wreck. In such scenarios Rand claims that what an agent ought to do, regardless of what might be the best consequences for an individual, is "to combat the disaster and restore normal conditions [that is,] metaphysically normal, normal in the nature of things, and appropriate to human existence" (Rand, 1964b, p. 54). This position has similarities with the two-factor account of flourishing developed in Chapter 8; it might be claimed that during emergencies the environmental factor (first criterion) of flourishing takes precedence and the individualist factor (second criteria) is put on hold until a situation is restored in which flourishing (for humans in this case) is once again possible.

It is common in applied ethics to examine moral dilemmas, that would (by Rand's definition) constitute emergencies, and to determine how an agent ought to act in these specific emergencies. An example of such a dilemma is Unger's (1996, p. 9) Shallow Pond scenario:

The path ... to the humanities lecture hall passes a shallow ornamental pond. On your way to give a lecture, you notice a small child has fallen in and is in danger of drowning. If you wade in and pull the child out, it will mean getting your clothes muddy" (Unger, 1996, p. 9)

Singer, Rachels, Unger and others have used this and similar scenarios to demonstrate that there is no moral difference between killing and letting die and that just as we *must* jump in to save the drowning child we *must* also help others in need, specifically those suffering from famine. If we must act to save the child (because the required sacrifice of clean clothes is trivial) then, by generalisation, where we can we must act to save those suffering from starvation. Acting to save someone from starvation is of great value (life) and requires a trivial sacrifice (a

⁹⁴ Rand prefers the term 'objectivist ethics' and very seldom uses the term 'egoism'.

few dollars) from those who live in absolute affluence (that is, those who can afford luxuries).

Let us take a closer look at Unger's scenario and for the moment assume that the girl in the pond is a complete stranger to you, no one else is near the lake and thus no one will witness you acting, whether that action be rescuing the girl or walking by and leaving the child to her fate. Further, let us assume that you have some specific interest in getting to this lecture on time (and in nice clean clothes). Many would argue that the consequentialist egoist, to maximise self-interest, ought to walk by. The egoist has an interest in being at the lecture on time and no interest in the complete stranger (and, being an egoist ought to have no pangs of conscience at having left the girl to her fate). Given that leaving a girl to drown solely to get to a lecture on time is (both intuitively and as a best explanation) a pernicious act such examples are usually given to show that consequentialist egoism is flawed. While I might quibble about what exactly the egoist ought to do in this specific scenario, as was demonstrated in Chapter 7, I agree that such scenarios appear to point to a serious flaw in consequentialist-based egoism⁹⁵.

Rand tries to navigate a middle ground between the altruistic argument that you must save the child and therefore have a similar duty to help those in need, and the extreme egoist position that you have no duty to help the child and in fact ought to walk by. Adopting the position that the passer-by *must* save the child (given that there is no danger to the passer-by) Rand once again demonstrates that it is not the specific consequences for an agent that dictate correct action. Her specific claim is that "If the person to be saved is a stranger, it is morally proper to save him only when the danger to one's own life is minimal; when the danger is great it would be immoral to attempt it" (Rand, 1964b, p. 52). Again, the egoist aspect comes through; the danger to oneself dictating whether the agent should act, but the egoism in question is not primarily consequentialist. The personal interests of the agent, which might be furthered by not saving the drowning stranger, are not given *any* consideration. Rand goes on to claim that the closer the drowning person is to the passer-by, emotionally speaking, the greater the risk the

⁹⁵ That is, egoism is dismissed, as a flawed theory, providing that a convincing theory of moral intuition is provided or my thesis regarding moral facts as best explanation is accepted.

passer-by ought to accept. The “greatness of the person’s value to oneself” (Rand, 1964b, p. 52) the greater the risk one ought to accept in attempting to rescue them. That is, the greater the value an agent puts on the person in danger the greater the risk the agent will be prepared to take in attempting to rescue them.

However, Rand denies that how we ought to act towards strangers in an emergency situation has implications for how we ought to act towards strangers in daily life. According to Rand (1964b, pp. 54-55), the fact that we ought to help strangers caught in a shipwreck does not mean that we are obliged to save these same people from “poverty, ignorance or neurosis” or “that we should sail the seven seas in search of shipwreck victims to save”. Rand (1964b, p. 55) goes on to point out that we “do not live in lifeboats”. That is, life does not consist of a series of emergencies. The vast majority of persons will never walk past a shallow pond and see a drowning child but will, on a daily basis, have to interact with strangers and as such the first scenario (shallow pond) represents an exception to the generalised rules of ethics, not the rule itself.

Rand advocates a partiality principle but exactly why one ought to save the child is unclear. While the following may not be a strictly Randian approach I suggest that the partiality involved is a limitation that is placed upon my first criterion of flourishing. A situation that inhibits flourishing in an agent’s immediate eco-system compels action in a way that flourishing in a distant eco-system may not. Any emergency that puts those in an agent’s immediate eco-system in danger requires an agent to act so as to normalise the scenario and tends to override individualistic concerns. For the time being it is enough to note that Rand’s theory is not, or at least is not primarily, consequentialist. The specific axioms of Rand’s theory are far more difficult to ascertain; due to Rand’s literary style it is not always clear what the premises and conclusions of her arguments are. Certainly Nozick (1971) and Den Uyl and Rasmussen (1978) disagree, both on the construction and interpretation of Rand’s theory. However, both interpretations together with Rand’s own work and Hunt’s (1999) neo-Randan thesis provide a platform from which to begin examination of a nonconsequentialist form of egoism.

§9.1 On the Randian Argument

Nozick and Rand share many of the same political (libertarian) views with regard to the rights of individuals, most notably, the belief in negative rights that prohibit the aggression of one upon another. However, whereas libertarians take these to be moral rights, for the most part when Rand talks of rights she doesn't use the term in a moral sense; these are legal rights, albeit legal rights that are supposedly determined or, as Rand claims, derived from her objectivist moral theory. I will not discuss these specific rights. It is difficult to fathom exactly how they are supposed to follow from Rand's moral theory and they are of little interest in regard to how Rand's ethical theory is formulated.

Before beginning an in-depth analysis of Rand's theory I want to make it clear that my reason for doing so is not to defend Rand's moral doctrine. I am doubtful a fully coherent theory exists within Rand's works, and indeed she doesn't claim to have a fully developed theory. Rather, her work consists of a series of ideas with regard to what she believes represents an ideal moral society. My goal over the next few chapters is to determine whether a nonconsequentialist theory of egoism can be developed using Rand's work, and the philosophers mentioned in the previous section provide a starting point.

§9.1.1 On the Randian Argument: Values

I will show that Nozick's reconstruction of Rand's objectivist argument is flawed in almost every respect. However, I consider it worth tracking Nozick's reconstruction as (i) it is by criticising Nozick that Den Uyl and Rasmusen more accurately reconstruct Rand's arguments, and (ii) Nozick's misinterpretation of Rand highlights some of the stranger positions attributed to her and, perhaps, the reason that her work is often overlooked. Nozick reconstructs Rand's initial premises as follows:

- (1) Only a living being is capable of choosing among alternatives, or,
- (2) Only for a living being could there be any *point* to choosing among alternative actions, for
- (3) Only a living being can be injured, damaged, have its welfare diminished, etc., and

- (4) Any rational preference pattern will be connected with the things mentioned in (3), and since
- (5) Values establish a (rational) preference ordering among alternative actions, therefore it follows that
- (6) Only a living being can have values with some point to them. Values have a purpose *only* for living beings (1971, p. 283)

Nozick's conclusion (6) seems strange in that, as Den Uyl and Rasmussen, (1978 p. 185) note, "Only a living being can have values with some point to them" is *not* the conclusion that Rand draws nor is it implied by any of her premises. Nozick's conclusion seems to suggest that values can exist apart from living persons (but without point), whereas Rand is not claiming "that only living being have values *with a purpose*" (Den Uyl and Rasmusen, 1978, p. 185), but rather that only living beings can have values *of any sort* (Rand, 1964a, p 15-16). For Rand the idea of intrinsic value is incoherent, values are instrumental to living beings. It is only the existence of living beings, beings that have the capacity to choose, that makes the concept of value possible. Perhaps a charitable interpretation of Nozick's conclusion brings it closer to Rand's position, that is if Nozick means that 'the concept of values would be pointless without living beings' but this is not what his conclusion state.

Nor does it appear that Nozick has correctly formulated the first premise of Rand's argument. Rand's claim is not that *only a living being is capable of choosing* but rather that *only man* has the capacity for choosing between the alternatives. For animals the code of values and choices between alternatives is automatic (Rand, 1964a, p. 20-21). As Den Uyl and Rasmusen (1978, p. 186) correctly note, Nozick has blurred Rand's distinction between "facing alternatives and choosing between them". According to Rand, animals face alternatives but the decision on which action to take is automatic whereas humans face alternatives and choose which path to follow.

Perhaps, if the principle of charity is applied, reasons can be found for Nozick's formulation, although the omission of any reasoning for his formulation makes this pure conjecture. It is possible that Nozick is simply attempting to formulate Rand's argument in its strongest possible form and in doing so has manipulated Rand's premise in order to overcome the objection that non-living things also face

(but do not choose between) alternatives. Any non-living thing faces the alternative of not being that thing any more. Whilst matter itself cannot be destroyed⁹⁶ a house can be demolished and while all the house parts remain the house itself ceases to exist. As Den Uyl and Rasmusen (1978, p. 186) recognise, there is a distinction between living beings and non-living things in as much as a non-living being does not “achieve or fail to achieve [goals, in this case survival] as a result of its actions”. However, the distinction that Den Uyl and Rasmusen make is not available to Rand who has already claimed that the actions of animals are automatic. If the actions of animals are automatic then animals do not achieve or fail to achieve goals as a result of chosen actions (nor do they set goals). They simply, according to Rand, act in a pre-determined manner towards a pre-determined goal. In this respect animals are no different from non-living things. Both animals and non-living things face alternatives and neither animals nor non-living things choose between those alternatives.

Perhaps realizing that Rand’s distinction between non-living things, animals, and humans was open to criticism, Nozick softened premise (1) to include animals (living things) and in doing so to make clearer the distinction between living beings and non-living things. Thus, premise (2) makes more sense in that the concept (point) of choice for a non-living thing is meaningless. Living beings (in some sense) choose between alternatives, non-living beings do not. The premise is of course still open to challenge as while it seems reasonable to ascribe choice to higher animals it seems equally unreasonable to ascribe choice to plants, despite being in the class of living beings. Nozick’s premise is simply wrong if it is supposed to be a direct interpretation of Rand, and flawed if it is supposed to present a stronger form of Rand’s argument.

Premise (3) ‘Only a living being can have its welfare diminished’ is extracted directly from Rand, the difficulty lies in determining exactly what role it plays in the overall argument. Den Uyl and Rasmusen, (1978, p. 187) criticise Nozick claiming that “The capacity to be harmed, damaged, or have one’s welfare diminished does not constitute the primary reason for claiming that either only

⁹⁶ At least according to commonly accepted scientific theory.

human beings choose or only living things act for one alternative as opposed to another”. However, it is far from clear that this *is* what Nozick is claiming by the inclusion of this premise. If Premises (3) and (4) are read together then the conclusion is not that ‘human/living things act in order to avoid harm, damage or diminished welfare’, but, as Rand herself claims, that the very recognition that some things are of value flows from the recognition that some things are good for us *qua* humans and some things will do us harm. Rand (1964a, p. 18) claims that “just as sensations are the first step in the development of a human consciousness in the realm of cognition, so they are its first step in the realm of evaluation” and in doing so infers the importance of the third premise. Premise (3) is vital in that it determines who can choose and *not* why we choose. Premise (3) sets the groundwork for determining who can have values, namely anything that can be harmed and has the capacity to be aware that it is being harmed. Which group of beings fall under this description is a difficult question; I would argue against Rand and claim that the group is wider than humans. However, I would agree with Rand that there is something in the distinction between those creatures that simply act and those with the capacity to make choices, but I am uncertain where and how to draw such a line. Certainly it excludes non-living things and equally certainly it is less than all living beings.⁹⁷ For the sake of convenience I will employ Singer’s (1993, p. 58) term ‘sentience’ as shorthand for meaning all those beings that fit under such a description of having the capacity to be harmed, damaged, or have their welfare diminished.

Nozick (1971, p. 284) does, mistakenly, assume that the values in question must, when acted upon, have some direct consequential benefit to the subject else they cannot actually be values at all. “I can’t value, or, in a valuing fashion, act to achieve some state of affairs, in a far off place, knowing that I shall never know whether my act has actually succeeded or not, and knowing that in either case its consequences will not affect me.” But this teleological aspect is simply not present in Rand’s determination of value. An act (one made based on values) must have an alternative and, in acting one way or the other, the act must make a

⁹⁷ By living I do not necessarily mean biologically alive. I accept the logical possibility of artificially intelligent robots that have the capacity to choose and might be considered alive. However, this difficult distinction and what exactly defines a living being is not something I will investigate to any great depth.

difference for the person acting, it does not matter if I achieve the goal of the act or even if I can ever know whether the goal has been achieved but that in acting I promote “those principles which determine the successful maintenance of one’s (i.e. qua human being) life or not.” (Den Uyl and Rasmusen, 1978, p. 188). If I have a value X and I act for that value X then I maintain that value by the very action itself, I need not ever know what the actual consequences of that action are. If I value honesty and in a given scenario am honest then I have acted for my value of honesty regardless of any consequences, of benefit or otherwise, that do or do not follow. It is the act itself and not the consequences of the act that determine if an agent has acted correctly, i.e. acted to promote the agent’s values.

What then might be drawn from the discussion of Rand’s early arguments with regard to value? I contend that:

1. Only sentient creatures have a welfare *and* are aware of that welfare.
2. It is only through the possibility of having a welfare and being aware of that welfare that values are developed (i.e. what is good for me).
3. Only sentient creatures have values.

I have not determined exactly what is included in the class of sentient creatures, other than the likelihood that it is larger than the set of humans and smaller than the set of all living things, nor do I intend to. My claim is conditional; if a creature is such that it has a welfare and it has the capacity to be aware of that welfare then that creature has the potential to develop values and, in some sense, to act on those values.

§9.1.2 On the Randian Argument: Life as Value

In Section II of Nozick’s article he attributes a second argument to Rand, concluding in the statement “life itself is a value to a living being which has it” (1971, p.283). Nozick’s (1973, p. 285) version of the argument runs:

- (1) Having values is itself a value
- (2) A necessary condition for a value is a value
- (3) Life is a necessary condition for having values. Therefore,
- (4) Life itself is a value

Nozick attacks this formulation of Rand’s argument on the grounds that premises (1) and (2) are false. This argument seems quite bizarre and, if it were the position Rand held would be cause for some concern. However, Rand does not hold either premises (1) or (2) nor does she argue for the particular conclusion that Nozick

claims. 'Life' for Rand is not just a value, it is the ultimate and fundamental value without which no other value exists.

An ultimate value is that final goal or end to which all lesser goals are the means – and it sets the standard by which all lesser goals are evaluated. An organism's life is its standard of value (Rand, 1964a, p. 17).

Metaphysically, Life is the only phenomenon that is an end in itself: A value gained and kept by a constant process of action (Rand, 1964a, p. 18).

The closest Nozick comes to capturing these concepts is in Premise (3) which, unfortunately, Nozick chooses not to discuss. However, as Den Uyl and Rasmusen (1978, p. 189) correctly determine, Premise (3) is "in fact her central ethical thesis". When Galt, Rand's hero/anti-hero from *Atlas Shrugged* is quizzed on the basis of his morality he replies "...in a single choice to live. Everything else proceeds from here.". Rand's argument is that life is the ultimate or fundamental value without which there can be no other values. Rand provides a transcendental argument in favor of the claim that life is the ultimate value, an argument that Den Uyl and Rasmusen examine in some depth and to which I will now turn my attention.

Rand (1964a, p. 16) criticises those who begin the development of ethical theories by asking which values are of importance and instead questions what values are and why we possess and/or need them. Rand's use of the term "value" is morally neutral; it simply means the object of an action, "that which one acts to gain and/or keep". As Den Uyl and Rasmusen, (1978, p. 189) infer, a value is the "goal" of a specific action. Given that we (I take 'we' to mean all sentient creatures) do appear to act, at least on occasion, in a goal-directed manner it can then be asked 'what are the conditions that make such goal-directed behaviors possible?' In reply to this question Den Uyl and Rasmusen, (1978, p. 189) extract three basic criteria from Rand's thesis:

- (1) Goal-directed behavior, by its very nature, implies that there is an alternative present.
- (2) Goal-directed behavior requires, by its very nature, the existence of an entity faced with an alternative.
- (3) Goal-directed behavior requires, by its very nature, that the alternative faced by an entity makes a difference.

Criterion (1) implies that if there are no alternatives, if we can act in manner A and only manner A then acting in manner A is not an example of goal directed behaviour, as no choice is being made. It is only because we could act in manner A or manner B and choose to act in either manner A or manner B that it can be claimed that choosing one of the two alternatives was (or at least could have been) goal-directed. When we must act in manner A and therefore do act in manner A it may well be that in so acting we promote a personal goal but that goal does not direct the behavior unless we have a choice, unless we could have acted otherwise.

Criterion (2) is self-explanatory. Goal-directed behavior requires the existence of some entity that can act in manner A or manner B. The alternative to act in either manner does not exist unless an entity faced with the alternative also exists.

Criterion (3) is the claim that, for an action to be goal-directed, acting in manner A or manner B must make a difference to the entity who acts. If it makes no difference to an entity whether that entity acts in one manner or the other then either no real alternative exists (both actions, though different, achieve the same end) or there is no difference for the entity between succeeding “or failing to achieve some goal, there would be nothing to differentiate [between] achieving and not achieving some goal [and] no alternative is faced by the entity” (Den Uyl and Rasmusen, 1978, p. 190).

As claimed in section §9.1.2, only sentient creatures have values and thus only sentient creatures have the capacity for goal directed behavior, a goal necessarily based on some type of evaluation. Only sentient creatures can face alternatives, act for a goal, and achieve or fail to attain that goal such that they are affected by that success or failure. What ultimately differentiates between success and failure, such that criterion (3) is met? According to Rand it is the fundamental difference between existence and non-existence. It is that an entity’s very existence is conditional, that an entity can choose between existing and not existing is the fundamental source of all the alternatives a sentient creature faces. Life is the ultimate and fundamental value and is implicitly valued by the making of any

other choice. To demonstrate this, Den Uyl and Rasmusen, (1978, p. 191) present the following argument:

- (1) X is an object of choice
- (2) Y is a necessary condition for the existence of X as a value. Y makes X as a value possible
- (3) If P chooses (values) X, P must choose (value) the necessary condition for P's valuation of X.
- (4) P chooses (values) X
- (5) *From 3,4* P chooses (values) Y
- (6) Y is life
- (7) *From 5,6* P chooses (values) life in choosing (valuing) X

Y can only be Life for the simple reason that for us to hold any X as a choice we must be alive. For any creature that can make the fundamental choice between life and non-existence the life of that creature (as the sort of thing it is) is that creature's ultimate value, and the standard of evaluation, for that creature. It simply makes no sense to claim otherwise; to state that I value X but do not value Y (life) even though Y is a precondition (and also a choice) for holding X is absurd.

Nozick's premise (3) "Life is a necessary condition for having values" is a central component of Rand's thesis, and Nozick's second premise is correct, if it is amended to read "A necessary condition for a value is an ultimate value", for which life is the only contender. As discussed at some length in §9.1.2, to claim that X is objectively superior to Y requires some evaluative criteria that has the property of superiority. In contrast to what I earlier contended, Rand's claim is that life is this value and that all evaluative decisions are ultimately derived from this.

The term 'life' without the conjoining "as the sort of thing it is" implies little specific meaning. By 'life' Rand is not referring to just the biological state of being alive but living in a certain manner (as is discussed more fully in §9.1.4). This being the case, the principles to be drawn from Rand's transcendental argument are as Den Uyl and Rasmusen, (1978, p. 198) conclude:

- (1) Life is an ultimate end, an end in itself, for any living thing.
- (2) To be a living thing and not be a living thing of a particular kind is impossible.

- (3) The particular kind of living thing that an entity is determines what one must mean when talking of “life” with respect to a given entity.
- (4) *Thus*, Life, as the kind of thing it is, is the ultimate value for each living thing.

My only divergence from this formulation is that ‘living things’ be replaced with ‘sentient creatures’, for the reasons discussed in the previous section.

§9.1.3 On the Randian Argument: Rational Persons

Nozick finds most of the arguments he has presented thus far, unsatisfactory. However, as I have shown, Nozick has failed to understand Rand’s basic principles. In section III of Nozick’s paper he moves on to consider Rand’s conclusion that “For each man, the preservation and prolongation of his life qua man, as a person, is a value for him.” (Nozick, 1973, p. 288). Again Nozick misses the main thrust of Rand’s arguments and instead focuses on rationality as being either (i) something that man and no other creature has (or can have) or (ii) as being the necessary essence of being human. Given that Rand does not argue for either of these positions I will disregard what Nozick has to say on them. It is only in the last paragraph of section III, as Den Uyl and Rasmussen (1978, p. 192) similarly note, that Nozick touches on the actual Randian argument. Unfortunately, other than this brief paragraph, Nozick fails to consider what is the most likely interpretation of Rand’s position. Thus I will begin, where Nozick chooses to finish:

It might be said that a rational person follows *principles*, general policies, and so we must consider those principles of action which make man’s survival possible. But it has not been shown why each person must follow the *same* principles, and why I may not as a rational being, have a clause in mine which recommends parasitism under certain conditions (Nozick, 1973, p. 292).

Value is that which we act for, the goal gained in choosing action A over action B. Given that it makes a difference for us (as sentient beings) whether we act in manner A or manner B then we are applying some form of evaluation and are acting, in Rand’s terms, to keep or gain some value. The ultimate value to any sentient creature, as a being of a certain kind (that is, being creatures of a certain nature), is the prolonging of life, as the type of thing it is for a creature of this nature. The principles and/or policies we apply in pursuit of these values are what

Rand refers to as virtues (though they differ somewhat from what is normally meant by this term, see §9.1.3.1)

Nozick rejects the claim that there are certain principles, policies or virtues that make human life possible as human life, on the grounds that it has not been shown that there are uniform principles that apply to all humans nor that a rational (human) being cannot work under a set of principles within a given context that allows or promotes parasitism. That is, according to Nozick, under a certain set of conditions a person could develop a set of values that allow that individual to act as a free-rider, to be a parasite upon society and live and benefit from the work of others. Rand would not deny that a rational individual could live as a parasite on humanity nor does her theory entail that everyone must adopt the exact same value set. What Rand is claiming is that there are a basic set of values that humans as a specific type of being must adopt, to live as humans. Humans, as a certain type of being, cannot live under a general principle of parasitism simply because humans (as a species) could not survive by such a principle. A general principle of parasitism would leave no society on which to be parasitic. The principles which Rand is referring to are not those that are simply logically possible, not those by which a single individual might survive, but those principles necessary for the survival of man qua man. Thus humans, as beings of the same kind, to survive as beings of that kind, must follow the same basic principles. This is not to say that an individual cannot act against these principles, but where we choose to do so, “we may, or may not live very long, but the fact remains that we would not be living the life of a human being; we would be a metaphysical misfit living by sheer luck and/or by the moral behavior of someone else.” (Den Uyl and Rasmusen, 1978, p. 193)

Rand would deny that a person living by a different set of values (such as parasitism) could be a successful one. Rand would counter that such a person would fail in the achievement of his own ultimate happiness. The approach is similar to the one extolled by Hursthouse (1999, p. 228) when considering if virtue ethics, based in ethical naturalism, can determine that a gangster is a bad human being. Her answer is yes given that “a gangster is bad qua human being”. A gangster, in being a gangster, possesses certain character traits (acts from

certain principles), which are the antithesis of those characteristics that have the best chance of promoting human life. Thereby a “gangster is bad qua human being, and thereby unable to live a good human life.” I am not wholly convinced by this argument nor of the claim that ‘good human life’ requires the adoption of a specific set of virtues or values, but I will not engage with the claim at this time.

For Rand, the goal of life is (for any specific being, as a being of that kind) to act in the manner that is suited to a being of that kind. For humans, that is to act in accordance with the nature of humans under the principles that will guide humans, as a being of a certain kind, towards successful lives as rational beings. Individuals can act against these principles but they cannot exclude themselves from these principles without also excluding themselves from the set of beings of that kind.

§9.1.3.1 On the Randian Argument: A brief outline of virtues.

Rand claims that, whereas values are what we act to gain or keep, virtues are the means by which we gain or keep them. Given such a claim it seems appropriate to pause at this stage, from extracting the specifics of Rand’s thesis on values, and consider exactly what Rand means by the term virtue. According to Rand the virtues are productiveness, rationality and pride. Rand (1964a, p. 29) describes the virtue of rationality as “the recognition and acceptance of reason as one’s only source of knowledge, one’s only judge of values and one’s only guide to action.” Rationality requires that one take responsibility for one’s own actions and that an agent is committed to gaining the greatest possible awareness of his/her surroundings and the fullest possible perception of reality. Every individual is responsible for acquiring the maximum possible knowledge (in any given context) and rationally acting upon that knowledge in accordance with the agent’s values. The virtue for rationality incorporates several other virtues, including; independence (accepting responsibility for one’s own life), integrity (not sacrificing one’s values to others), honesty (accepting reality and not skewering it in order to justify some goal realised out of desire rather than rationality), and, justice (that one accepts only what is deserved and likewise bestows only upon the deserving (remembering that this applies only within what might be considered normal life, not to emergencies).

The virtue of productiveness is “the recognition of the fact that productive work is the process by which man’s mind sustains his life” (Rand, 1964a, p. 29) Productive work is the rational pursuit of some worthwhile (valued) result. The virtue of productiveness requires that the mind is engaged and used in a rational manner and that an individual not act like a zombie, simply moving through life as a matter of routine.

The final (primary) virtue is pride by which Rand means the acquisition of “values of character that make life worth sustaining – that as man is a being of self-made wealth, so he is a being of self-made soul.” (*Atlas Shrugged*, cited in Rand, 1964, p. 29) Pride then is the development of a set of values, of which one’s own life is the highest value, and the recognition by an agent that the agent’s life, as a certain type of thing, is the agent’s highest value and that the agent has a right to pursue that value.

I introduce this material only so it is clear what Rand means when talking of virtuous action to gain or keep one’s values. I will leave discussion of which virtues can be derived through a naturalistic process until the following chapter.

§9.1.4 On the Randian Argument: Social Rights

The depth of Nozick’s misconception regarding Rand’s position becomes clear in Section IV when he presents the three-step argument:

- (1) For each person, the living and prolongation of his own life is a value *for him*
- (1’) For each person, the living and prolongation of his own life (as a rational being) is the *greatest value* for him.
- (2) No person should sacrifice his life for another
- (3) No person should sacrifice another person (’s life) for himself (his own.)

This argument cannot be attributed to Rand. Firstly, Rand does not claim that a person’s life is the *greatest value* for him. To refer to a *greatest value* in this manner infers that life is a greater value than some other set of values when measured against some criterion. Rand’s claim is that life is the ultimate value; that which makes all other values possible. Perhaps Nozick meant exactly this but

it seems strange at this point to drop Rand's terminology of life as an ultimate value and replace it with the term "greatest value". Secondly, by 'life' Rand is not referring to the prolongation of biological survival. As Rand (1964a, p. 26) makes explicit:

It [life for rational beings] does not mean a momentary or merely physical survival. It does not mean the momentary physical survival of a mindless brute, waiting for another brute to crush his skull. It does not mean the momentary physical survival of a crawling aggregate of muscles who is willing to accept any terms, obey any thug and surrender any values, for the sake of what is known as "survival at any price" which may or may not last a week or a year.

It is life, as a certain kind of thing, which is the ultimate value. Rand does not call for the prolongation of biological life at the expense of sacrificing those rationally held values that were necessary to an agent's life, as a rational being, being a life worth living. The prolongation of one's biological life is not the ultimate value and Rand provides two examples where one might sacrifice biological life because not to do so would require an agent to sacrifice that which, as a rational agent, defines that agent's life as a life of value. The first example comes from the ethics of emergencies where Rand (1964b, p. 52) claims that if the person one loves is in mortal jeopardy then one ought to be willing to risk one's life to save them, as to not do so would make life unbearable. The principle (virtue) of integrity requires being loyal to one's values, (and assuming the term love is being appropriately used then that love is of the highest value). Thus if a person does not act, where rescue is a possibility, despite any risk associated with the rescue, that person is failing to act in accordance with that agent's own values and is, in effect, sacrificing those values. It is this exact type of sacrifice (sacrificing a greater value, love and a life worth living, for a lower value, biological survival) that Rand labels immoral. Another example, though unrealistic, comes from *Atlas Shrugged*, in which Galt considers suicide rather than be forced to live in a world devoid of what he considers to be of value.

Nozick, in formulating his argument, seems to ignore the fact that what Rand is interested in are the guiding principles for the survival of rational beings as beings of that kind and it has already been demonstrated that Rand would deny that 'individual survival at any cost' is a suitable principle. The ultimate principle, for

Rand, is not that no person should sacrifice his (biological) life for another but rather that one ought not sacrifice a greater value for a lesser value, “any action that a person undertakes for the benefit of those he loves is not a sacrifice if, in the hierarchy of his values, in the total context of the choices open to him, it achieves that which is of greatest personal (and rational) importance to him” (Rand, 1964b, p. 51).

To be fair it is easy to find reasons why Nozick may have constructed the argument in the manner he has. Rand does state early in her thesis that the fundamental alternative living things face is the choice between existence and non-existence and it is from this that everything else follows, a claim that would seem to imply biological survival. However, as has been noted on several occasions, Rand was not writing as an academic philosopher so that much of what she says must be taken in the overall context and when this is done it is clear that Rand is referring to far more than biological survival (and certainly more than individual biological survival). However, simply taking Rand’s comments in context does not alleviate all of the problems in interpreting her work. It is often difficult to determine if by the term ‘man’ she is referring to human beings as a type of being or to an individual agent. However, putting these difficulties aside, Den Uyl and Rasmusen’s analysis of Rand’s overall project seems far more accurate or, at the very least, is presented in a style which more accurately reflects the ideals and concepts that Rand extolled.

The remainder of Nozick’s argument attempts to analyse the move from values to individual rights. This analysis suffers greatly from the failure to correctly interpret the principles of Rand’s moral argument. Rights, for Rand, involve the implementation of values within a social context. While she does sometimes refer to these as moral rights (although for the most part simply political or legal rights) they are not part of the basis of Rand’s morality in themselves. I will therefore leave Nozick’s argument at this point as nothing additional can be gained from examining whether social rights can be derived from values. I will briefly turn my attention to Rand’s concept of happiness before returning to further consideration of what is implied by Rand’s constrained egoism. The egoist principle is that I ought to act so as not to sacrifice (or in positive terms, to promote) ‘my’ values

where those values are determined by myself, as a rational agent. The constraints are that values are restricted by what it is to be a rational human agent and that individualist values are overridden where the situation is such that the very possibility of flourishing requires acting in a certain manner.

§9.1.5 On the Randian Argument: Happiness and the Good of others

For Rand, humans, as a creature of that kind, ought to act from rationality. The ultimate goal for humans is happiness which can only be gained from having lived a life with the values appropriate to a being of a certain kind. Rand does not claim that happiness will be obtained by living such a life, merely that it is only through rational pursuits (for rational creatures) that true happiness can be obtained. “Happiness in Rand’s philosophy functions much the same way as in Plato’s and Aristotle’s. It is something that takes time and training to achieve.” Happiness is not some instant gratification or pleasure, rather “happiness is a lasting and continuous sense of rightness with the world” (Den Uyl and Rasmusen, 1978, p. 201). This concept of happiness puts considerable constraints on Rand’s egoism. An individual’s happiness cannot be separated from ‘the way the world is’ and ‘the way the world is’ to a considerable degree depends on the actions of others and interactions with others in order to attain that state of rightness with the world.

This may at first appear to be some form of enlightened hedonistic egoism and that somehow the well-being of others brings about one’s own well-being. I contend this is not the case. Rather, Rand’s concept of rightness in the world has more to do with a world in which certain values can be promoted. Hunt (1999, p. 83) appears to interpret Rand in a similar fashion and in describing an agent’s interests claims “one’s own interest [and thereby, ultimately, happiness] is (consists in) the attainment of value, and one of the most valuable things is the good of others – that is, certain other – people.” As has been claimed throughout this chapter, Rand is not taking a consequentialist approach. It is not that bringing about good for others results in a better life for me. It is simply that the attainment of an agent’s values will include (as part of those values) the good of (some) others. Hunt (1999, p. 81) claims that, as such, a person acting on that individual agent’s values (i.e. acting to gain or keep those values) embraces “the common-

sense view that the good of others is a ground floor reason for action in that it is worth pursuing in itself". While I think Hunt is on the right track it might be objected that he has quietly dropped the *some* from his conclusion; the conclusion that follows from Hunt's remarks is that the "good of some others is worth pursuing" and that the good of others is not worth pursuing in itself, it is worth pursuing because not to do so would be irrational, it would be to act in breach of the agent's own set of values. The common-sense view that Rand would embrace is that the good of some others is worth pursuing because to do so is to promote one's own values and that those values provide a ground floor reason for action. The values are worth pursuing in themselves.

I have previously noted (see §9.1.4) Rand's claim that to love someone obligates an agent, in acting for that value (love), to act for the good of the other. Rand has further comments regarding how we ought to act toward those we call friends, and to those who are strangers. Rand's claim regarding friends progresses along the following lines; values are rationally derived from what it is for a human to exist as a creature of a certain nature. It must be allowed that humans are, by nature, social creatures and that for rational beings friendships form a vital part of being social. Thus humans, as social creatures, require friends and the friendship (not the friends themselves) will constitute a value. To say that someone is a friend, and to value friendship, is also to acknowledge that keeping that value will require acting for the welfare of that other (providing that doing so does not result in the sacrifice of a greater value). If a person claims some other is a friend and that friend faces difficulties then that person is compelled to help them (by the person's own set of values) except where doing so would mean the sacrifice of a greater value. What would be of greater value so that a person ought not to help a friend in need? This question is difficult to answer; Rand never explains how the values are supposed to be placed within a hierarchy (other than noting the process would be a rational one based upon the ultimate value of life). Presumably love outweighs friendship but this would be little help in working out a detailed application of the theory. However, that a theory is difficult to apply is just that, a difficulty: it does not point to a flaw in the theory itself unless the difficulty is one that cannot be overcome. I believe the difficulty can be overcome and that developing a personal hierarchy of values is unproblematic and on most occasions

agents will have no problem in ranking values. However, it will on occasion be extremely difficult, if not impossible, to differentiate between two values. Such occasions are simply an accurate reflection of moral choice; moral dilemmas can arise, situations where no matter what action an agent takes they will be forced to sacrifice one of two equal values. I will return to the problem associated with determining and ranking values in the next chapter.

When it comes to how an agent ought to treat strangers (individuals who are unlikely, even indirectly, to figure in an agent's hierarchical set of values) Rand (1964b, p. 53) has this to say; we ought to grant strangers "the generalised good will which one should grant to a human being in the name of the potential value he represents – until and unless he forfeits it". According to Rand we ought to respect all other human beings as beings of the same form of life. Again it is difficult to determine exactly what the value is (perhaps respect) and where it lies within the hierarchy of values, certainly less than friends and loved ones and, it would seem, certainly more than trivial interests. It would seem that a value of respect for human beings would require that, except in emergencies where we are compelled to help, how we ought to act toward strangers entails simple non-interference as per the principle of 'not sacrificing another to oneself'.

§9.2 Constrained Egoism

It might reasonably be argued, at this point, that Rand's doctrine is a form of constrained egoism (action is determined by self-interest except in certain contexts) and that such a theory is a farce; the real ethical theory is not the pseudo one being described but the one that determines the constraints. The constraint in this case is the notion that an agent cannot sacrifice others even if doing so is to the agent's benefit, a constraint that appears decidedly non-egoistic.

It is true that there are many cases where the good of another is a constituent part of our own good. If a loved one fails at some endeavour then an individual may suffer because he values that other, or more specifically values the relationship. However, assuming such an egoistic view of the value of friendships works, that still leaves a vast number of other cases where the good or otherwise of another will have no impact whatsoever on the agent considering how to act. It is not only

that in many cases an agent will be able to predict (with a reasonable probability of success) that the other in question will never represent a value to the agent. Why then, if the agent stands to gain substantially, not sacrifice the stranger in order to make that gain? Why should Rand's egoist have respect for all other human beings purely because they are human beings? The answer Rand gives seems to rely on the agent's actions being restrained by a doctrine of flourishing; namely what it is for a human qua human to flourish.

In attempting to interpret Rand's theory as a form of flourishing-based egoism Hunt (1999, p. 89) agrees with Rand in that my friend can be a part of my good, claiming: "If I do take this position there is no *prima facie* reason to think that I will advance my interests by stealing his wallet, even if he never suspects me and, in purely consequentialist terms, I 'get away with it'. If my relationship with my friend is, to use Rand's terminology, 'one of my highest values', then by betraying his trust and victimizing him I would be damaging my own life just as I am damaging his". Indeed, this seems a reasonable response from a nonconsequentialist egoist with regard to not stealing a friend's wallet but how can such a defence work with regard to strangers? Someone I have never met and am unlikely to ever have a relationship with is unlikely to hold a particularly high rank in my hierarchy of values. Why then shouldn't I steal the stranger's wallet? The reason why is that it is difficult for humans to flourish where theft is commonplace and that flourishing of the community requires that I constrain what might appear to be a self-interested action (though I am dubious, outside of a controlled thought experiment, that stealing a wallet is ever likely to be in an agent's true self-interest). Hunt (1999, pp. 92-93) argues that such a position is "equivalent to describing a way of life" a set of values under which humans will flourish. Hunt (1999, p. 93) goes on to argue that such a way of explaining self-interest will lead to the determination of "why certain character traits have the status of virtues".

If it can be shown that by the adoption of certain character traits (and the valuing of these traits) humans give themselves the best chance of flourishing then we may find that acting with a certain manner towards strangers is indeed in an individual's self-interest. It is to this task that I will turn my attention in the next

chapter, namely, can the virtues (as specific values or character traits) be derived from an agent's self-interest in flourishing and being happy (assuming, as I will, that humans qua humans desire to flourish and be happy). However, before attempting this task I will briefly introduce one criticism of flourishing-based egoism.

§9.3 A brief Criticism of Randian/Flourishing Egoism

My purpose in this chapter is to grapple with the basic premises of the Randian argument and the elementary ideas of an egoism based on flourishing. My intention has not been to defend Rand or claim her theory is particularly solid or coherent, in fact I agree that there are many flaws in the theory which invite criticism. One such criticism can be demonstrated by altering slightly Hunt's (1999, p. 80) scenario of the man and his dying wife:

Suppose a man's wife is deeply ill and it would cost the man's last \$10,000 to save her. The man used to love this woman but has long since fallen out of love with her. Now suppose the man also has a company that employs 10 people, friends he cares about deeply. Injecting the \$10,000 into the company would save the company from bankruptcy and thereby also increase his friends' welfare etc.

It would seem by Rand's theory that the man ought to invest the \$10,000 into the business, increasing his friends' welfare and thereby keeping his greater value. His wife should be left to die, certainly a sacrifice but a lesser sacrifice than (a) allowing his business (which he values) to fail, and (b) sacrificing the welfare of his friends by seeing them unemployed. Such a position feels counter-intuitive and it seems that saving his wife's life (regardless of his feelings towards her) is more important than saving his business and his friends' unemployment, (although, as I have earlier stated, I put no great faith in intuitions and perhaps mine are simply mistaken in this case). However, it does begin to highlight the extreme difficulty in actually applying Rand's system. How exactly is the hierarchy of values supposed to be determined? Nor does it appear to provide any way to deal with states of conflict. Rand's only answer to the latter is that states of conflict do not exist amongst rational men.

Rand's denial of conflicts between rational men is demonstrated using something of a false analogy; the conflict she dissolves never really exists. Rand's example

involves two people applying for the same job and she argues that no conflict of interest exists between the job applicants primarily because rationality requires that both men realise only one can have the job. The very existence of the job requires a commercial environment, and excellence in business requires that more than one applicant be available for a specific job etc (1964c, p. 64). This may well be true enough, but this is insufficient to show that conflicts do not exist. In considering my earlier example (§2.5) of the four poisoned lab technicians this would seem to represent a very definite conflict of interests. In order to pursue any of the agent's values, they necessarily must get the antidote. The recognition that only one can get the antidote only serves to heighten the conflict. I have earlier claimed (§2.5) that there are irresolvable conflicts and that no ethical theory can provide the means by which to resolve these conflicts. However, these occurrences are not the norm and it need not be a requirement of an ethical theory that it can solve each and every state of conflict.

§9.4 Conclusion

The neo-Randian notion of value examined in this chapter expands on the individualistic factor of flourishing; an agent flourishes when that agent is not forced to sacrifice that agent's values, or is not forced to sacrifice a greater value for a lesser value. Rand also touches on certain objective values (virtues) that appear to be derived from an objective notion of eco-flourishing and what it is for a human to flourish qua human beings.

In Chapter 10 I will expand on these ideas and demonstrate that a hierarchy can be developed that is partially holistic and partially individualistic. A combination of virtues that humans must adopt if humans are to have the best chance of flourishing together with individual (subjective) values that determine what it is for an individual human to flourish.

Chapter 10 Flourishing Egoism

§10.0 Introduction

In Part 1 of this thesis it was determined that if consequentialist egoism is a consistent moral theory then the good must be relative both to an agent and that agent's worldview (§3.4). Further it was noted that an egoist would, in most cases, be motivated to act in order to promote that good. However, the nature of good was left undefined. Having since discussed flourishing (Chapter 8) and Rand's thesis of acting so as not to sacrifice one's own values (Chapter 9) I will now defend a version of ethical egoism based on these principles. I will define the good as being that which an agent values, which necessarily (for any rational agent) involves that agent's flourishing.

The complete principles of flourishing- or value-based ethical egoism are:

- (i) *(Definition of Value)*
 Ψ_x is a value to an agent if;
(a) Promoting or holding Ψ_x is most likely to lead to or bring about an agent's flourishing, as a certain kind of being, relative to *any* rational worldview,
or
(b) Promoting or holding Ψ_x is constitutive of what it is for an individual agent to flourish relative to that agent's worldview.

- (ii) *(Universal principle of Ethical Egoism)*
For any n, if Ψ_1 is of value to agent_n relative to agent_n's worldview
and
If Φ ing would promote agent_n's value Ψ_1
and
If Φ ing would not require the sacrifice of something in which agent_n instils greater value (Ψ_2)
then
Acting so as to promote the value Ψ_1 provides agent_n with strong moral reason to Φ , and justification for Φ ing.

- (iii) *(Individual motivation for acting)*
If agent_n believes that ethical egoism provides preferable reasons for acting then agent_n will, in most cases, be motivated to act in accordance with (ii).

For the most part this chapter will involve a discussion of principle (i), the definition of value, and the difference between (a) the values that any rational agent must hold if that agent is to flourish as a kind of being, and (b) subjective values that are of importance only to a specific agent, and are a constitutive part of an individual's flourishing. I will develop these principles using the discussed notions of flourishing (Chapter 8) together with the neo-Randian principle of an agent acting so as to promote the agent's own values (or at least acting so as to not sacrifice greater values for lesser ones).

§10.1 Beings of a Certain Kind and Universal Values

Hursthouse claims that the adoption of certain character traits (her example is of a gangster who is dishonest, ruthless, greedy etc.), while not necessarily precluding that agent's flourishing, is less likely to result in that agent's flourishing than would the adoption of traits more commonly assumed to be virtuous in nature (i.e. honesty, benevolence, kindness). Such a position appears, intuitively, to be correct. However, on closer examination the claim quickly loses much of its intuitive force. It is true that modernised societies tend (as a rule) to shun criminals and wherever possible take punitive actions against those who break societal laws, resulting in a society where it is difficult for criminals to flourish. However, this same modernised society embraces capitalism and multinationals that are ruthless (e.g. hostile takeovers), greedy (the driving force behind corporations is increasing profits and rising share prices), and often dishonest (or at least not honest, e.g. the deliberate withholding of information preventing customers from making informed decisions) and these businesses *do* flourish. The flourishing of a business and human flourishing are clearly not the same, but these business enterprises *are* run by people and in many cases are run by a single person. Given that individuals are directly responsible for the methods adopted by the business enterprises they control, it would seem that to some extent the individuals themselves are adopting, or at least promoting, these negative character traits. While I do not want to claim these individuals are flourishing, it seems odd to suggest that a vast number of individuals in western society are involved in practices, and adopting traits, that prevent them from flourishing;

traits that they could fairly easily change⁹⁸. Ultimately I agree with Hursthouse; honesty is more likely to result in flourishing than dishonesty. However, at first sight Hursthouse's thesis is unconvincing given the vast numbers of people who appear to flourish by the adoption of negative character traits.

In an argument that bears some resemblance to Hursthouse's, Rand claims that being a rational human being excludes certain types of behaviour (the holding of certain values and the adoption of certain character traits) on the grounds that to act in such a manner would (i) be to act irrationally, and (ii) adopting such behaviour as a maxim would be to exclude oneself from the class of humans (as a certain type of being). In short, to act in a manner destructive to flourishing (for a certain kind of being) is to exclude oneself from the class of beings of that kind. The statement is difficult to refute; if all Rand is saying is that it is irrational to adopt certain character traits that prevent the agent from flourishing, where rational action is defined as acting in a manner that leads to flourishing, then the claim is trivially true. All Rand has done is define rationality and state that those who don't act rationally (so defined) are irrational. At best Rand would be making an existence claim, namely that there are character traits whose adoption prevents an agent from flourishing.

Hursthouse does outline those virtues most likely to result in human flourishing but this is mostly done by negation (e.g. the gangster analogy) and, as noted above, such a claim is not compelling. Further, as discussed in Chapter 8, I think her naturalist project is flawed and with it the conclusion that specific character traits are likely to result in human flourishing; if the groundwork that leads to a specific claim is flawed then doubt must be cast over the conclusion drawn from that reasoning. I am not denying that the virtues Hursthouse extols may lead to an agent's flourishing, only that (1) it has not been shown that the extolled virtues are more likely to result in an agent's flourishing, and (2) the reasoning behind extolling these virtues is flawed.

⁹⁸ I do not mean to suggest that capitalism is evil or in some way necessarily results in companies that are dishonest, greedy etc. My aim is simply to show that this does happen and companies that are as a rule dishonest, greedy, and ruthless do flourish. Numerous documented cases support this, a recent and highly documented case being the McLibel Trial (See: McLibel, *Two Worlds Collide*, 2004, Madman)

Rand bases her theory in self-interest (the individual pursuit of a life worth living) and suggests the character traits (values) we ought to adopt are directly related to our individual self-interest. However, Rand is extremely vague in outlining which values we ought to accept, how those values are to be determined, or how a hierarchy of such values is to be formed. Other than Galt's speech from *Atlas Shrugged* and the very general claim that life is our ultimate value and everything else flows from that, we are presented with no clear indication of what a Randian hierarchy of values might look like. Applying the principle of charity, and a reasonable degree of interpretation, the 'Randian value system' can be extracted from Rand's discussion of happiness. According to Rand an agent obtains *real happiness* only when *things are right in the world* (from that agent's particular perspective). I interpret '*things being right in the world*' to mean something like:

- (i) Reality_s is such that an agent's flourishing is logically possible
- (ii) It is actually within the agent's control to flourish, and
- (iii) Such flourishing makes the agent's life, as a certain kind of being, worth living.

For Rand, 'life' is an agent's ultimate value; a complex value that incorporates the entire set of an agent's subjective values and determines the nature of a 'life worth living' for an agent as a 'kind of being'. This ultimate value will consist of many subjective (agent-relative) values but it also places considerable restraints on what an agent can value. Agents, as rational beings, cannot value that which is destructive to the flourishing of the species and still claim to be a rational member of that class of beings. For example, while an agent could hold a set of values based on death and destruction (perhaps claiming to value war) in doing so it would seem that the agent's own flourishing is precluded, limited, or at least put at serious risk and that, counterintuitively, what the agent values is ultimately the agent's own self-destruction. It might be countered that there are those who do thrive in conditions of war and do indeed flourish under these conditions – I am doubtful this is true. Certainly some persons can profit from war and many even seem to enjoy the state of conflict, my doubt is that these people are actually flourishing. A rational moral agent, in this case an ethical egoist, must accept the principle that individual flourishing is dependent on the world (Reality_s) being a place where individual flourishing is a possibility and that the ethical egoist has

strong reason to ensure, where it is within the agent's power to do so, that the world is such a place.

Even allowing that some can flourish in such conditions as a state of conflict, such flourishing is limited. They can flourish *only* if wars are limited in scope; a world engulfed in endless war would soon see the end of any flourishing. A rational agent cannot hold death, destruction and war as values as, if it is accepted that a value is something an agent will promote wherever it is possible to do so, then promoting the agent's values of death and destruction ultimately, and irrationally, leads to the detriment of the agent who holds those values. If the agent claims to value a world beset by endless conflict the result of which will be an end to flourishing then the agent is being inconsistent; the agent desires an end to that which makes holding the value a possibility. If the agent claims to value war (elsewhere) and peace (where the agent spends the wealth gained from limited warfare) then again irrationality appears, the agent holds conflicting values (the agent cannot promote the values of both war and peace). The agent might claim to value a state of limited conflict but I am doubtful this is possible, more likely the agent would be acting as a consequentialist egoist promoting war where such conflict was of financial benefit and promoting peace where it was not.

Such examples do little to provide any specific answers or bring us any closer to understanding what constraints might be placed on an egoist's actions. However, they do demonstrate the means by which to do so. All that needs to be determined is what it means for the world (Reality_s) to be such a place that the agent can flourish, in order to determine that actions detrimental to this state of the world are also detrimental to, or likely to negatively impact on, the agent.

§10.1.1 An Agent's Worldview and Holistic Values

Principle (i.a) *Definition of Value* states that Ψ_x is a value of an agent *if* promoting or holding Ψ_x is most likely to lead to or bring about that agent's flourishing as a certain kind of being and further that Ψ_x will be a value regardless of the worldview the agent holds, so long as that worldview is rational.

The flourishing reference in the above principle can be either holistic or objective. By holistic flourishing (see §8.2.2) I am referring to the biotic community, the

community that the rational egoist must admit he is dependent on and therefore values. By objective value I am referring to those values that an egoist must hold in order to have the greatest probability of flourishing regardless of the specifics of the agent's worldview. I will discuss objective value in §10.1.2 but first will consider what limitations might restrict an egoist worldview.

While some advanced or final science might prove otherwise, I am assuming that there is *not* some specific and objective state of the world X that can be pointed to or described and said to be the perfect flourishing ecosystem. The flourishing of an ecosystem is multiply realisable. Thus when I use the term 'rational worldview' what I mean is that an agent, as a certain kind of being, holds a rational worldview *if* the class of beings of which the agent is a member is likely to flourish (and the ecosystem flourishes in a manner that will support that class of beings). This does little to inform us about how such a worldview might be constituted but it does tell us a great deal about what will not constitute such a worldview.

As discussed (§10.1) an agent cannot rationally seek to create a world engulfed by war; the holistic system upon which the egoist is dependent would collapse and with it the egoist's ability to flourish. Thus, while the egoist need not necessarily value peace and tranquillity the egoist cannot value a state of anarchy and destruction. At the very least the egoist must value an absence of unnecessary conflict and destruction⁹⁹. Similar conclusions can be reached with regard to the egoist's attitudes (and ultimately values) towards health; a rational being could not flourish in a world ravaged by disease and while the egoist may not value an end to all disease the egoist must value a reduction in disease, controls on virulent viruses etc. If disease is not controlled then humans will not flourish and in all probability the egoist will likewise fail to flourish. An egoist will value controls on diseases, as the egoist's own flourishing requires minimising the possibility of contracting virulent disease. Similarly, while an egoist need not value an end to world hunger the egoist cannot value famine for, in a world ravaged by famine, beings that require food might manage to survive but they will not flourish. The egoist's value is that there is sufficient food such that the possibility of flourishing

⁹⁹ As demonstrated in §2.5 sometimes conflict is unavoidable.

is present. I contend that such values will rank highly in the egoist's hierarchy of values (as they are necessary for the very possibility of flourishing), but these values are preferential. The egoist will more highly value an absence of famine for himself than an absence of famine for others.

The egoist will value air clean enough to breathe (the degree of pollution an egoist might deem acceptable being dependent on the full set of values held by the egoist), biodiversity (to the degree necessary to support a food chain/biotic pyramid of some sort), forests (again for air), drinkable water etc. For an egoist to do otherwise, to value that which would render water undrinkable, the world plunged into war, the atmosphere rendered unbreathable would be to will the world was such a place (or to hold such to be a value) is to will that beings of the same kind as the agent will in most cases fail to flourish and counterintuitively, that the agent holding the value will most likely fail to flourish. In short, if the egoist holds certain negative values then (against the doctrine of egoism that I have begun to outline) the egoist is valuing that which is detrimental to eco-flourishing, the egoist's own flourishing, and ultimately to a life worth living.

One criticism of the view so far, briefly touched on in §10.1, is that an agent might hold a worldview where that specific agent can and indeed would flourish but where all, or rather most, other agents suffer. Further, it would seem, given the egoist framework, that no call can be made to a veil of ignorance or, in this case, universality; worldviews *are* relative to an agent. Such a worldview might be possible for a few (dictators for example) who have direct control over the lives of others. In a dictatorship the person in power knows they have this power (and may have been born into such power e.g. the Caesars of Rome) and has the ability to create a world where they flourish at the expense of others. However, I am doubtful that such a worldview is truly viable and that a rational agent could claim it is the most probable way by which that agent can flourish. It would seem that the only way for the downtrodden to flourish is to overthrow the dictator and in this respect the dictator places himself in a precarious position, necessitating constant guard against revolution, assassination, or internal coup. If the dictator must be on constant guard then the flourishing that can be experienced is already

limited. Further the dictator's flourishing could disappear at any moment and be replaced by 'a life of misery in jail' or 'violent death'.

Very few of us have the power of a dictator or the ability to *actually* shape our world. Further, it is unlikely that the dictator holds any of the negative values outlined above. Rather I suggest the death and destruction that is wielded is more likely a consequence of other (perhaps misguided) values on the very fringes of what might be deemed rational. For example, taking an extremely simplified view of history, Hitler would not be said to have valued a world plunged into war, rather his values of a restructured Germany overrode his values for peace and security and resulted in a scenario where, for Hitler, not to go to war would have been to sacrifice a greater value. I am not suggesting that Hitler was ethically correct in going to war only that if his worldview was rational and his value system well formed (both of which are unlikely) then Hitler would have acted correctly in promoting his greatest value.

To recapitulate, an egoist must place value in an ecosystem that allows beings of like kind to the egoist the possibility of flourishing; a flourishing ecosystem is essential to the egoist's health, security and general well-being. While the specifics of a valued ecosystem are subjective, the fact that the ecosystem is valued is universal. Exactly what these values are is difficult to determine given that a flourishing and sustainable ecosystem is multi-realizable. Further, a rational agent will value their own flourishing and given that an agent can only flourish in a context where others flourish it would be irrational to promote a situation in which agents tend not to flourish.

So far I have given indications of the sorts of values an egoist clearly cannot hold. I will now turn my attention to objective values and specific character traits that an egoist must hold if the egoist is to act in a manner more probable to lead to an agent's flourishing. I gave an indication of one such value in §2.4.3 and §6.9, namely 'Honesty'. Both MacIntyre and Rogers noted the importance of honesty, not for the good of others, not as a duty to others, but because of the benefit to the individual. I will now examine the claim in more detail.

§10.1.2 An Agent's Worldview and the Importance of Honesty

We (i.e. any specific agent) could adopt a random set of virtues and see if we flourished while acting in accordance with that set of character traits. Then we could go on to modify the set of virtues and repeat the process over and over, each time determining the degree to which we flourished under the given set of traits, and eventually determining which traits led to the maximum degree of flourishing. However, while such a task is theoretically possible it is not practical. Firstly, virtues are things that make up part of our being and are not something we can change at whim. Secondly, testing the results and comparing degrees of flourishing would not be easy. Thirdly, the process would likely be self-defeating; one's entire life could be spent testing sets of virtues at the expense of the very goal of the project, an agent's flourishing. If an agent is to develop certain character traits in the belief that these traits are most likely to result in the agent's flourishing, then the reasons for adopting those virtues must be clear and compelling to that specific agent. The most compelling reason I can conceive of for an agent to adopt certain virtues, values or character traits is that acting in a certain manner is in the agent's self-interest and is the most probable means by which the agent can ensure that they will flourish.

The flourishing of any agent regardless of the specifics of that agent's worldview requires that the agent can have faith in the reality of the worldview they have adopted. To recapitulate my claims of Chapter 6: to flourish we need to be able to make sense of the world and making sense of the world requires appealing to best explanations. The best explanation of what would most probably lead to my flourishing requires self-knowledge, as "self-knowledge is necessary, if I am to imagine realistically the alternative futures between which I must choose" (McIntyre, 2001, p. 95). True self-knowledge requires an absence of self-deception. At the very least we must be honest and truthful with ourselves if we are to flourish, but not only must we be honest with ourselves. If we are to trust our senses (and we must do so if we are to build up a picture of the world) then we must honestly express what we see to others and, knowing that others (or at least most others) will act likewise (if they are to have any faith in their worldviews), receive confirmation of our perceptions. Only by this ongoing interaction with others can the egoist build up a picture of the world in which the

egoist can feel secure. My claim is that for an agent to have a reasonable probability of flourishing that agent must establish faith in a worldview. Further, given that faith in a worldview requires confirmation (firstly via the absence of self-deception, secondly via others) the egoist has good reason to develop the character trait of honesty. That is, an ethical egoist will place a high degree of value on honesty such that developing the character trait is in the egoist's self-interest. An agent who has not adopted honesty as a character trait, instead using honesty/dishonesty as a tool and acting in the manner the agent believed best suited the situation would be unable to place much trust in others (believing others ought to act in a similar manner), would be unable to find reassurance for the agent's own worldviews, and would soon begin to lose faith in any facts. It is not that a dishonest person could not flourish; rather the claim is that an honest person is more probable to flourish and therefore it is in the individual's self-interest to develop the character trait of honesty.

I have not defined specifically what I mean by honesty. For the most part, in the discussion above, I have used being honest as the opposite of deliberately misinforming or lying. Dishonesty could also be defined in terms of withholding the truth, white lies, and non-verbal deception. My claim is that all of these would be contrary to the egoist's values but that honesty/dishonesty exists on a continuum and certain forms of dishonesty are low on the continuum (e.g. white lies, or not supplying unsolicited truths) and are more likely to be tolerated simply because being honest on these occasions is more likely to result in the sacrifice of greater values. For example, not withholding honest opinion on someone's attire, when the truth is likely to result in a widening chasm between yourself and someone important to you, is to place too high a value on this particular form of honesty.

A critic might respond that this is a strange form of egoism given that it makes continual reference to others (and to the fact that the agent's honesty benefits others). However, the egoist is *not* being honest for the good of others or to promote a functioning society, although these are obvious offshoots. Rather, the egoist values honesty, and is honest, primarily because it is necessary for establishing self-knowledge and rational worldviews.

- (i) I have an interest in developing a character trait of honesty (to avoid self-deception).
- (ii) I have an interest in others being honest (to provide confirmation of worldviews).
- (iii) I assume that others are (generally) honest for the same reasons (i.e. it is in their self-interest to be honest).

To say that the egoist values honesty, as with any other value, is to say that the egoist will promote, care for, and preserve that value. It makes little sense to say I care for X but I choose not to act in a manner that will preserve X (given no other contributing factors). This does not mean that the egoist will always act honestly (as with the white lie example), or that 'Be honest' is an absolute rule (or in fact a rule at all), only that the egoist will develop a character trait of honesty and will tend to act honestly without conscious consideration. The egoist will not treat honesty lightly and unless some greater value would be sacrificed by an honest act, has no reason to act dishonestly; even then the dishonest act will be out of character and require mental effort.

Rand claims that virtues are the means by which one keeps one's values (see §9.1.3.1). To a degree I agree with Rand but I think she misses the essential point; namely that while certain virtues and/or character traits are the best means by which to ensure an agent flourishes (taking 'an agent's flourishing' to be roughly equivalent to Rand's use of the term 'a life worth living as a certain type of thing') the agent will necessarily value those things that enable the agent's flourishing. An egoist values honesty because, given any reasonable view of the way the world is, it is highly probable that being an honest person is more likely to result in the agent's flourishing than if the agent were to treat honesty as a commodity to be used or discarded as the agent saw fit.

That honesty increases the probability of an agent's flourishing is compelling reason for an agent to value honesty and adopt honesty as a character trait. Unless an agent is placed in a situation where acting honestly would require the sacrifice of something in which the agent instils even greater value, then the egoist will not even consider a dishonest act. Flourishing-egoism differs little from other moral theories in its promotion of honesty but (to put it somewhat crudely) whereas a utilitarian is honest only in as much as honesty generates the greatest good or

maximises preferences, and a deontologist is honest as a matter of rule, the egoist (in a manner similar to virtue theory) is honest because it leads to and is constitutive of the agent's own flourishing.

§10.1.3 An Agent's Worldview and the Importance of Respect for Others

In order for an agent to flourish he/she needs to feel secure. Someone in constant fear of attack or of losing one's possessions cannot flourish (or at least any flourishing is extremely limited). An individual in such circumstances has two options. The first is to live behind a wall of security and bodyguards, and while not an option for the majority a few individuals are in such a position (dictators, the extremely rich etc). However, if history has taught us anything it has shown us that such positions are anything but secure. Bodyguards are corruptible, armies overthrow leaders, and even the highest level of security can be of very little defence against a fanatic. While such measures may lead to an agent feeling secure, I doubt that such a person can flourish given the restraints such security measures would place on the individual. It would seem that a far better method of obtaining a feeling of personal security is through removing, or greatly reducing, the reasons others might have to breach your personal boundaries. By allowing for and taking account of the boundaries of others, that is by respecting the personal space and property of others, an agent reduces the likelihood of, and a motive for, others to invade the agent's space. It is in an individual's interest to respect the space of others and in doing so not to provide others with a reason to invade the individual's space. Likewise, it is in the interest of others not to invade a specific individual's space if they are aware (and it is in an agent's interest to make them aware) that doing so will likely result in retribution. Such situations were discussed in §2.5 where I demonstrated that in most cases a consequentialist egoist has no problem with conflict resolution;

Clearly it is not the case that egoism is unable to solve any conflicts of interest. For example, if some resource (R) exists and both Agent₁ and Agent₂ have an interest in R then while a conflict of interest appears to be present it is not the case that a Darwinian struggle need necessarily follow. For either agent to monopolise R is not in that agent's self-interest given that both agents also have a self-interest in avoiding unnecessary conflict, conflict that would likely follow if one agent was to monopolise the resource.

Likewise, the flourishing egoist will, in most cases, avoid conflict. However, I am not suggesting that the egoist ought to respect others, and gain respect from others, as a matter of reciprocal interest. Such a position is problematic because not everyone is capable of harming an individual and such an approach would lead to an egoist only respecting some other based on the degree of harm they could do to the egoist. Moreover, this is to slip back into a consequentialist frame of thought. Rather I am suggesting that because the flourishing egoist will highly value personal security the egoist will value character traits that best promote this value and that developing the character trait of respect, that is respect for the space and property of others, will directly result in a feeling of security for the egoist.

What egoism compels in this case is little different from how most people do act. We take care not to invade the space of others lest we face repercussions, and we expect others not to invade our space lest we take retributive action (or request others take such action on our behalf). Further, we do not invade the space of those who have little if any chance of retributive action because our actions are a matter of character not conscious consideration of consequences. At the same time we take reasonable precautionary measures (such as locking the house) to protect our interests.

The notion of respect for others outlined here differs greatly from the traditional moral notion that some inherent value of others makes them worthy of respect. This is to be expected from within an egoist framework; however, I maintain that what is described in this section is respect for others, if the reason for that respect is steeped in self-interest. Of greatest importance to an egoist is the pursuit of those things that the egoist values and in respecting the space of others an egoist is respecting the self-interest of others to pursue their values free from interference; in effect respecting others even if the reason for doing so is one of self-interest.

§10.1.4 An Agent's Worldview and the Importance of Limited Benevolence

Many will find it counterintuitive, if not blatantly contradictory, to mention benevolence and egoism in the same sentence. However, I will demonstrate that a limited form of benevolence is essential to an agent's self-interest; if an agent is selfish then certain systems, which the egoist values, are harmed. If those systems

that the egoist values are to have the highest probability of thriving then the egoist must develop an attitude of limited benevolence.

Consider, by way of analogy, an extremely simplified capitalist system, but one not unlike that which is common in western society. The basic goals of this system are the expansion of capital and the pursuit of profit, yet these goals lead to monopolisation (as by way of capital expansion the successful envelops the less successful), which in turn ultimately leads to the collapse of the capitalist system/worldview. The collapse of such a capitalist system can, in part, be prevented by government-imposed restrictions that restrict capital growth and limit the absorption of competitors. However, the thriving of such a system requires that those who are most successful within capitalist society and value the capitalist system act benevolently and inject money back into the capitalist system (even where there is no direct benefit) in order to ensure the stability of the capitalist system.

Using a rather crude example, Bill Gates' recent (November 2002) donation of \$100 million to fight Aids in India is of no immediate financial benefit to Gates. However, the injection of \$100 million back into society will filter through to the wage packets of many individuals who, in spending those wage packets, will perpetuate the capitalist system, a system Gates benefits from (and presumably values). Thus through donating \$100 million, in an act of limited benevolence, Gates promotes and cares for something he values, namely the capitalist system. I am not attempting to pass judgement on capitalism nor am I suggesting that the given example encompasses all the relevant details. I am merely attempting to show that protecting and promoting a system that an agent values may require what superficially appear to be unrelated acts of limited benevolence. Gates' donation of \$100 million is an act of benevolence yet at the same time it is a self-interested act, motivated by the desire to ensure the stability of a capitalist system and the worldview that embraces this system. Under this analysis, assuming Gates is both an egoist and values the capitalist system, the underlying reason for the act of benevolence is not a desire to help cure Aids in India but a realisation that the continuation of the valued system requires injecting a percentage of

absorbed capital back into society, and thus the agent who values capitalism has an interest in developing a character trait of limited benevolence.

By 'system' I mean the way that any group functions; it need not have a formal set of rules. For example, the family unit is a system and its members interact as part of that system. Likewise clubs, the work place, sporting events are all systems of interaction. An agent who values 'the family' will act with benevolence towards that system of interaction, that is the agent will act to ensure that the system of value (in this case the family) is sustained. This may be by way of caring for the individuals within the system (sharing the wage packet, providing adequate nutrition, ensuring access to education etc.) or by acting in a manner to promote the system itself (recognition of the family unit). If an agent values a church parish, a club, a sports team, or any community then that agent will take a benevolent approach towards that community; it is in an agent's self-interest to care for and promote those systems that the agent values.

This needs a little further clarification. I am not claiming that an egoist ought to weigh the pros and cons of a benevolent act and act only if it is of benefit to the egoist; this would once again be to slip into a consequentialist-based egoism. I am claiming that benevolence is a character trait that (regardless of worldviews) the ethical egoist needs to develop in order to flourish. However, a principle of partiality is in place; the egoist's benevolence will be tempered and restricted by those systems that the egoist values. To not act benevolently, to fail to support a system the egoist values (unless such an action would require the sacrifice of a greater value), even if the harm done to the system may be trivial, is to act against the dictates of flourishing egoism.

§10.1.5 An Agent's Worldview and the Requirement of Temperance

In order to flourish an egoist must exercise self-control. The need for temperance will become clearer when the subjective values are discussed (§10.3). However, as temperance is an objective value for the egoist I briefly introduce it at this stage. The unguarded pursuit of that which the egoist subjectively values can, in itself, be detrimental to that value which the egoist pursues (or other values that the egoist holds) and thus the egoist needs to develop a trait of temperance, and in

this regard values temperance. To offer a brief example, an egoist may value fine foods. However, the untempered pursuit of fine foods may lead to overindulgence which in turn is likely to affect the egoist's health, where good health is essential if the egoist is to flourish.

§10.2 Returning to Rand – Formalising the Principles

Making use of the basic Randian objectivist argument and adapting it to take account of the notion of flourishing, the position I am developing is:

1. Only sentient creatures can flourish *and* are aware of that flourishing.
2. It is only through the possibility of flourishing and being aware of that flourishing that values are developed (i.e. what is good for me)
3. Only sentient creatures have values.

Simply put, it is because we can flourish and realise that some things work for that flourishing and others are to its detriment that we develop a system of values. (See §9.1.1) It was further argued (§9.1.2.1) that

4. Life is an ultimate end, an end in itself, for any living thing.
5. To be a living thing and not be a living thing of a particular kind is impossible.
6. The particular kind of living thing that an entity is determines what one must mean when talking of "life" with respect to a given entity.
7. *Thus*, Life, as the kind of thing it is, is the ultimate value for each living thing.

To have a good life, to flourish in a manner appropriate for a rational sentient being is our ultimate value, it is the standard from which all other values are derived and placed within our personal hierarchy of values. In §9.1.3.1 I discussed Rand's claim that values are what we act to gain or keep, virtues are the means by which we gain or keep them. I have developed this into the claim:

- 7.i Certain objective values can be determined from accepting that personal flourishing is the ultimate standard of value for rational sentient creatures.
- 7.ii Developing certain virtues or character traits provides the most realistic possibility of gaining and keeping these values.

The specific claims outlined in §10.1.1 through §10.1.3 were that if *my flourishing* is *my ultimate value*, as it must be if I am to count myself among the set of rational beings, then my personal flourishing requires that (1) I have certainty with regard to my surroundings so that I can, with certainty, define Reality_s, (2) I feel secure within my surroundings, and (3) that those systems which I value, and

thus constitute part of my flourishing, also flourish. I then argued that the most probable way of gaining certainty and security within Reality_s and ensuring those systems that I value flourish was through the development of the character traits of honesty, respect, temperance, and (limited) benevolence. In Rand's terms, through the development of honesty and respect an agent will secure the values of certainty and security which are necessary to having a life worth living as a rational being. In my terms an egoist will come to highly value honesty and respect as a result of valuing certainty and security and an agent will have the best possibility of personal flourishing through the development of the virtues of honesty and respect. Further, the egoist will be benevolent (in a limited sense) as an agent's flourishing is tied to the flourishing of those systems the agent values. Those systems that the agent values are partly constitutive of the agent's own flourishing; an agent cannot flourish if those systems the agent values are left to fail.

The egoist will value honesty and respect; trivially this honesty is towards others and the respect is for others. Honesty and respect gain their importance for the egoist not through being other-regarding but as necessary character traits that are constituents of the egoist's flourishing. Further, the egoist will be motivated to perform benevolent acts (in the limited sense) not for the good of other individuals (although, this necessarily follows), but for the good of specific systems that the agent values.

To continue formalising the principles of Ethical Egoism:

8. Flourishing requires certainty with regard to Reality_s.
9. Certainty with regard to Reality_s requires the character trait of Honesty.
10. Flourishing requires security within Reality_s.
11. Security within Reality_s requires the character trait of Respect.
12. Flourishing within Reality_s requires that those systems that are constitutive of Reality_s also flourish.
13. Promoting those systems that the agent values requires taking a benevolent approach towards those systems which requires the character trait of (limited) Benevolence.
14. Acting in a manner such that the pursuit of values does not, in itself, become detrimental to those values requires acting with temperance.

My claim so far is that an ethical egoist must possess the character traits of Respect, Temperance and Honesty regardless of the worldview the egoist holds.

The egoist need not always be honest nor always respectful but will value these traits highly and will not lightly violate these principles. Again I stress that these virtues are not adopted out of reciprocal interests or mutual co-operation with other agents. Rather, the egoist adopts these virtues because they offer the best probability of the agent's own flourishing and, in promoting these virtues, the agent acts out of self-interest regardless of the specific consequences of the act. However, as noted above, the egoist will not always be honest; the egoist will sacrifice this value when not to do so would be to sacrifice a greater value (the hierarchy of values is discussed in the following sections). I further claimed that the egoist must develop a character trait of benevolence, again regardless of the worldview, but that this benevolence would be limited. The limitations placed on benevolence are subjective to the agent and more specifically the agent's specific worldview, and it is to these subjective values that I will now turn my attention.

§10.3 Subjective Values

I take X to have subjective value if an agent cherishes X and would feel a loss if X was in some way damaged. These values, much like the virtues, are enfolded in and constitutive of the agent's character. Values, again like the virtues, are developed over time and are not things adopted on a whim. However, values differ from the virtues in that whereas temperance, honesty, respectfulness and (limited) benevolence are objectively valuable (due to their essential nature in an agent's flourishing) and cherished because they are valuable, all other values are dependent on an agent and the agent's specific worldview. Despite value being subjective, this is not to say that an agent can value absolutely anything. Rationality and flourishing, while not determining what is of value, place limits on what those values might be. I have already shown (§10.1) that an ethical egoist cannot value that which will result in the agent failing to flourish; while an individual may hold such values they are inconsistent with the notion of flourishing and the holder of such values would not be an ethical egoist.

As to what the egoist can subjectively value, the list is endless. Anything that the egoist takes a stance towards, that the egoist cares for and would act so as to protect or promote is a value for the egoist. For example, an egoist might value sentient life, in any form, and consider such life to be precious (and I am

assuming, for convenience sake, that no other values come into conflict). Such an egoist would develop an attitude towards sentient life and, perhaps, become vegetarian or vegan, support certain charities, and attempt to persuade others (who hold differing values) that they were mistaken in their attitudes towards sentient life. The egoist will be aware that such values are subjective and that no absolute claims can be made. However, the egoist can, in this case, best promote this value by trying to convince others that the values they hold are misplaced, so long as in doing so the egoist does not sacrifice any higher values. Whether such a sacrifice occurs depends upon how highly this particular value is placed within the egoist's hierarchy. It could be that the egoist more highly values respect for others, in which case it would be wrong for the egoist to press someone with regard to their attitude towards sentient creatures such that the value of respect was sacrificed. However, if the egoist places respect for all sentient creatures as amongst the highest of values then, even though the egoist is normally respectful of others, to remain silent while others acted in a manner that caused damage to something upon which the egoist placed great value would be unthinkable. All things being equal, an egoist who valued all sentient life could not remain inactive while others tortured an animal, and continue to claim that all sentient life was valued¹⁰⁰.

I have given no real indication of how the values might be ranked for the simple reason it is not possible to do so. An agent's hierarchy of values is subjective. The objective values will, as they are necessary to an agent's flourishing, rank very highly but they are not overriding as my example re the value of sentient life demonstrates.

That a hierarchy of values is subjective leads to certain apparent problems, not least of which is how a hierarchy of value might be ordered. Obviously an agent need not have a formal list of values but there will be times when an agent must choose, sacrificing one value in order to promote another. In most cases this will present little problem as, from a personal standpoint, some things will have much greater value than others and the sacrifice of the lesser value will occur without

¹⁰⁰ While I will not press such a line of argument here it might also be argued that accepting life as an ultimate value obligates an egoist to value all sentient life not just life for one's own kind. While the value placed on the life of a cow may be a very distant value it may still outweigh such trivial values as taste.

much conscious deliberation. Dilemmas will occur when the egoist must sacrifice one of two values, both of which have great value to the egoist. Egoism provides no easy answer to such dilemmas. However, this is a strength rather than a weakness; it would be odd, when facing a ‘Sophie’s Choice’ and being forced to choose between one of your two children, to suggest a moral theory must give a clear answer. Whenever an egoist faces such a choice, that one of two near inseparable values must be sacrificed, whatever action the egoist takes their will be moral remainder; the egoist is left with ‘dirty hands’ and regret in having been placed in a situation where such a sacrifice was necessary.

A further apparent problem for egoism is that agents cannot objectively judge the actions of others. I do not consider this to be a real problem. Flourishing- or value-based egoism is a moral theory in that it guides the actions of the individual who ascribes to it; it is not designed to be judgemental of the actions of others (except in as much as whether or not they meet the mandates of the doctrine of egoism). While flourishing-based egoism requires the development of certain character traits (such as honesty, respectfulness, temperance etc) in that these traits provide the egoist with the greatest possibility of flourishing, no mandate of egoism is broken by being dishonest if by acting honestly the agent would sacrifice some greater value.

It should be noted that the reason an egoist cannot judge others is purely empirical. It would be logically possible for an egoist to make such an objective judgement if the egoist had knowledge of the values held by others and the weight others placed on those values within their personal hierarchies. However, that in practice an egoist cannot make moral judgements does not prevent the egoist from discussing the wrongness of specific actions or from challenging others’ actions. Firstly, the worldview and values of others can be discussed and critiqued and if the holder of such values cannot justify them then the perception of right action in the given context may come into doubt. Secondly, we can question how a specific action could possibly lead to an agent’s flourishing, and possibly convince them that an alternative action is preferable and has greater potential for the agent’s flourishing. Thirdly, as I noted in the discussion of moral facts (Chapter 6) there are limitations placed on what can constitute a rational worldview, and that right

and wrong is relative to the agent's worldview. The limitations placed on what can constitute a rational worldview results in certain (extreme) actions that are close to being objectively wrong. For example, even though I have no direct access to X's worldview, if X is torturing an innocent I can be reasonably sure that X is acting wrongly given the improbability of a scenario where the torturing of an innocent was the course of action that was most likely to result in that person's flourishing.

§10.4 Initial Objections

I imagine one of the first objections would be that the theory I am extolling is not really egoism at all but rather some sort of confused virtue ethics that, while formally egoistic (the virtues benefiting the agent that holds them), the reason for holding the virtues is to be a good person, not to be a good person for the agent's own benefit (which may or may not follow from acting virtuously). This is something of a reverse of the attack often made against virtue ethics; namely that virtue ethics is a form of egoism. However, the underlying premise is that an egoist ought to act so as to protect that which the egoist values. Given that the rational ethical egoist must value the agent's own flourishing I have claimed that the agent must also value certainty (§10.1.2) and security (§10.1.3) and that the best means by which to protect these values is by adopting the character traits of honesty and respectfulness. The only character traits I have claimed an ethical egoist must adopt are honesty, temperance, respectfulness and (a limited) benevolence because these are essential to an agent's flourishing. The adoption of any other so-called virtue is dependent on the specific agent's worldview and the values that may arise out of that worldview. Despite the arguments presented here, it is theoretically possible that acting so as to promote an agent's own flourishing might require the adoption of character traits commonly assumed to be vices. I am extremely doubtful this would actually be the case (I can see no way that an egoist could flourish in a world populated by dishonest, disrespectful people), but if it can be shown to be otherwise I simply bite the bullet and say, yes, the egoist ought to adopt those traits regardless of the assumption that such traits are vices.

It might be claimed that my conclusion that an egoist must adopt certain character traits is based on reasoning that is not egoistic; rather the theory is a disguised

form of utilitarianism. If everyone (or most people) were honest then I could gain certainty with regard to my worldview (as could others) and everyone would be better off if people were honest. This is not the case; the egoist is honest, first and foremost, to avoid self-deception. Secondly, the egoist is honest in portraying the egoist's worldview to others in that it is logically impossible for the egoist to receive confirmation of a worldview unless the egoist portrays such a view accurately. The egoist assumes others will follow the same logic and likewise (in most cases) be honest¹⁰¹. It is certainly true that everyone is better off if the majority of people are honest and that this benefits the egoist, but the primary reason for adopting the character trait is the agent's interest in the agent's personal flourishing. That everyone is better off if most people are honest may be true but it is not the primary reason for the egoist's honesty. Nor is it the case that the egoist is honest, and expects honesty from others, out of reciprocal interest; the egoist is honest out of self-interest and expects others will be honest out of their self-interest. The same is true of respect for others. While everyone (and therefore the egoist) will likely be better off if the majority of people are respectful of each other; the egoist's reason for being respectful of others is that if the egoist lacks respect for others the egoist is likely to face repercussions of some form. However, this is not a case of co-operating with others simply to avoid repercussions (at best such an argument would ensure the egoist was respectful of some others), rather the egoist has the best probability of flourishing if the egoist develops the character trait of respect for others.

A further objection (already briefly discussed) might be that the theory seems somewhat arbitrary and offers no clear action guidance in that no clear hierarchy of values exists. For example, the egoist will be honest unless by being honest some greater value would be sacrificed. But how is it determined which personal values are significant to the degree that dishonesty is justified? Faced with a situation where an honest answer will place a business in jeopardy (which the egoist has spent years developing and highly values), ought the egoist answer honestly or dishonestly? This criticism only appears difficult to respond to if there

¹⁰¹ At the very least Agent_x can draw the assumption that in a given scenario Agent_y is unlikely to suffer from telling the truth and is therefore unlikely to put other values at risk by attempting to deceive Agent_x. While this offers no guarantees Agent_x can reasonably infer that it is probable that Agent_y is being honest.

is an assumption that such questions should have a clear answer. My response is that it is one of the strengths of ethical egoism that it gives no distinct answer. In such situations the egoist faces a moral dilemma; the egoist must either sacrifice the value placed in honesty or risk sacrificing the value placed in the successful business. The solution to the dilemma would rest on the degree of dishonesty required (and the ongoing untruths that might be necessary following an initial untruth), the degree of sacrifice the egoist would face by the loss of the business, and the degree of risk that the business might be lost. Flourishing-based egoism offers the means to examine such dilemmas; it does not provide an absolute rule on how to act.

§10.4.1 An example of Ethical Egoism in Action: The Lifeboat Dilemma

Returning to the lifeboat scenario (§1.1) where the lucky egoist is the only one aboard a lifeboat that can support up to a maximum of 15 persons at an increasing level of risk. How would the ethical egoist approach this dilemma?

1. The ethical egoist recognises that the egoist's own flourishing is the egoist's ultimate value and (in this case) can be equated with survival. The egoist will act to ensure the egoist's own position in the lifeboat is safe.
2. The egoist values honesty (and in this case avoiding self-deception), and must recognise that she can save four passengers at no additional risk to herself.
3. The egoist in valuing respect must recognise that by not saving any (not respecting the others' lives) she sacrifices that value and must live with that sacrifice. Secondly, she recognises that the sacrifice of this value may lead to repercussions initially from those in the sea trying to board the boat and then from those on land (as discussed in §1.1) seeking recompense for their loss of value (i.e. the loss of family and friends whom they value).
4. [from 2 & 3] The egoist will save four passengers at no risk, reducing the risk of repercussions and, with more hands to staff the boat, increasing the chances of her own survival.
5. [from 1,2,3] The egoist will continue saving people until the sacrifice (the potential risk to her own life) exceeds sacrifice of respect for others (and accepting the risk of repercussions)

The exact number of passengers the egoist will save is difficult to determine but it seems probable the balance will be similar to the one reached by the virtue ethicist, attempting to balance the foolhardy act of trying to save everyone (and

thereby sacrificing the egoist's own flourishing) and the cowardly act of saving the minimum number of passengers (sacrificing respect for others and putting the egoist's flourishing at risk from the repercussions of others). Again this may sound like consequentialist reasoning, but it is not. The egoist does not consciously weigh the outcome in terms of risk to the egoist either by drowning or by repercussions (although the egoist will be aware of these considerations). Rather, the egoist, having necessarily developed a character trait of respect for others, could not consider leaving others to drown except where a greater sacrifice was called for. Thus the egoist would not initially even consider the evaluation process given above until it neared the point where acting out of respect for others would require a foolhardy sacrifice of the egoist's greatest value, the egoist's own flourishing.

§10.4.2 Ethical Egoism in Action: Rachels' Doctor

It seems fitting, given that I claimed consequential egoism fails as a moral theory in that it compels an agent toward pernicious acts, to subject flourishing egoism to the same test and investigate how Rachels' doctor would act if he was a flourishing egoist.

First we need to consider what set of values the doctor might have embodied. Perhaps the doctor was an extreme capitalist and placed enormous value in the acquisition of wealth. If this were the case then the doctor would have developed character traits that would have given him the highest probability of attaining such wealth in his chosen career; certainly honesty (as a dishonest doctor is unlikely to have an ongoing clientele), certainly a degree of respect for others (for much the same reason). Given these character traits the doctor would not even consider charging for a service (an examination) that was neither given nor necessary. Perhaps then the doctor valued wealth to such an extent he would sacrifice the values of honesty and respect in order to gain an extra twelve dollars? Such a claim seems improbable, twelve dollars is a trivial increase in wealth (to a doctor) and is extremely unlikely to justify the sacrifice of two virtues that are essential to the doctor flourishing. I can find no set of values that the doctor might embody that would lead to the doctor taking the woman's last twelve dollars. The Doctor might, as a flourishing egoist, have refused to see a patient that could not afford to

pay (which would not require the sacrifice of any values) but, while harsh, it is unlikely such an act could be labelled pernicious.

Chapter 11

A Day In Punxsutawney

§11.0 Introduction

In this final chapter I will demonstrate how the axioms of flourishing- or value-based egoism might be applied by way of an extended case study. The case study I will use is the one presented in the movie *Groundhog Day* (1993). The original script for *Groundhog Day* was steeped in Zen Buddhism and is often supposed to demonstrate the progression from an egoistic or Nietzschean nihilistic position towards the discovery of a good life through virtuous living. Such a position is developed by Kupfer who claims that by the end of the film Phil (the central character) no longer sees “other people as opportunities for his personal advantage, he now views himself as a resource for the benefit of others” and, rather than “demean people, Phil goes out of his way to spare people’s feelings, including sparing them the embarrassment of being grateful to him” (2005, p. 293). I will argue that Kupfer is mistaken and that while Phil realises he cannot simply use others as opportunities for personal gain he does *not* see himself as a resource and does *not* sacrifice himself, or his values, for the benefit of others. I will further argue that Phil’s progression is not from egoism towards virtuous living but rather is a progression from consequentialist egoism towards a flourishing-based ethical egoism. Phil’s journey is the personal discovery of self worth and the development of values through an understanding of true self-interest and not, as Kupfer would claim, the recognition of some inherent good in others (2005, p. 293).

§11.1 A Day In Punxsutawney: The Scenario

Phil Connors, a selfish, egocentric, and thoroughly dislikeable weatherman, is assigned the task of reporting the Groundhog Day festivities in Punxsutawney, Pennsylvania, for the fourth year running. Phil is bitter about having to travel to Punxsutawney and has nothing good to say about the trip, the town, or the people who reside there and intends to leave at the earliest possible moment. However, an unexpected blizzard prevents his departure and he is forced to spend the night in Punxsutawney. Phil awakens to the sound of his radio alarm at 6am, to discover that it is still Groundhog Day. We are given no reason for this bizarre

phenomenon, nor is one really required, the scenario is simply that no matter what Phil does, he will continue to wake at 6am each morning and it will always be Groundhog Day. Phil is trapped inside a repeating 24-hour loop. However, the scenario is further complicated in that only Phil has memories of the ongoing repetitions, everyone else believes they are experiencing this Groundhog Day for the first time. As Phil soon realises, if tomorrow never arrives then he has no reason to worry about long-term consequences:

[Phil] What if there was no tomorrow?

[Bar Patron] No tomorrow? That means there'd be no consequences, no hangovers, we could do whatever we wanted.

[Phil] That's true we could do whatever we want. (*Groundhog Day*, 30:20)

And the stage is set, similar to the story of Gyges (§2.3), for Phil to ponder the question 'Why be moral?' given that Phil does not have "to live by their rules anymore!" (*GD*, 32:25). Whatever action Phil takes there will be no lasting consequences, at 6am everything will be reset and no-one will be aware of any acts that Phil committed on the previous day.

Despite the fantasy setting of *Groundhog Day* the movie is analogous to many everyday lives; at one point Phil queries the bar patrons "What would you do if you were stuck in one place and every day was exactly the same, and nothing you did mattered?" the reply being "that about sums it up for me" (*GD*, 29:11). Thus Phil is pondering not just how he ought to act within this scenario but how he ought to act period. Rammis, the Director of *Groundhog Day*, notes that in searching for answers Phil progresses through five stages that closely follow Elisabeth Kübler-Ross' five stages of death and dying; Denial, Anger, Bargaining, Depression, and Acceptance. (*GD*, Commentary, 32:25) I will discuss these five stages, focusing on bargaining and acceptance, as these are more relevant to the moral progression from consequentialist- to flourishing-egoism. While in the bargaining stage Phil is at his most egocentric, acting purely to satisfy immediate desires, whereas in the acceptance stage Phil acts benevolently. Kupfer argues this is out of a general concern for others, I will show that it is actually limited benevolence necessary to Phil's flourishing and the promotion of self-interested values.

§11.2 Denial and Anger

Phil's initial response is disbelief and denial, thinking that the day simply cannot be repeating. He continually questions people and sets little tests, such as breaking a pencil the night before. However, everything is reset the next morning, the pencil is once more restored and everyone else is convinced they are experiencing this year's *Groundhog Day* for the first time. No longer able to deny the day is repeating, Phil's disbelief turns to anger and he begins taking his frustrations out on whoever is at hand. It is only when Phil realises that he can do whatever he pleases without any risk of repercussions that we get to witness Phil's egoistic hedonism in full flight.

§11.3 Bargaining

The section's title is a little misleading; Phil's actions are more a case of manipulation than bargaining. Without the threat of undesirable consequences to temper his actions Phil "goes to work grabbing for all the pleasurable experiences he can cram into the ever present *Groundhog Day*" (Kupfer, 2005, p. 286). He eats whatever he pleases, smokes as much as he wants, drinks to excess, steals money, and manipulates those around him for pleasure and personal gain. Phil's unique situation allows him to easily compile information about an individual that can be put to mischievous use in a future (repeat) occurrence of *Groundhog Day*. For example, Phil gathers information about a woman, her name, where she grew up, her English schoolteacher, so that he might pretend to be an old school friend in order to advance a desired sexual relationship. Yet, despite Phil's unique advantage over the other residents of Punxsutawney the one thing that Phil really desires, a relationship with his producer Rita, seems beyond his grasp. Rita recognises Phil's actions as selfish and his friendship insincere, chastising him:

The wretch centred all in self
living, he shall forfeit fair renown
and doubly dying, shall go down
to the vile dust from whence he sprang
unwept, unhonoured and unsung. (*GD*, 35:30)¹⁰²

Still, Phil continues his attempts to seduce Rita, using each revolution of *Groundhog Day* in his plan to create the perfect date. At one stage it seems Phil is

¹⁰² The quotation is the closing lines from a poem by Sir Walter Scott

close to success; acting in the manner he believes Rita will approve of, Phil's actions appear to be genuinely virtuous in nature. However, when Phil claims to love Rita she perceives the ruse and ends the date with a slap to Phil's face. The more Phil tries to control the outcome the earlier Rita recognises the ruse and, in response to Phil's professed love, responds:

[Rita] I could never love you because you'll never love anyone but yourself.

[Phil] That's not true I don't even like myself. (*GD*, 53:00)

Phil is unable to find happiness through a life of egoistic hedonism and, perhaps recognizing the truth in his own words, that there is nothing to like in his character, begins his slump into depression. Phil cannot have that which he really wants (a relationship with Rita) and, while he can satisfy almost any desire, Phil tires of this life of pleasure and soon becomes "bored beyond endurance" (Kupfer, 2005, p. 287).

§11.4 Depression

Bored with immediate pleasures, Phil becomes depressed with his life and enters his nihilistic stage; he sees no value in anything nor the possibility of anything having value, and kills himself over and over again. However, suicide offers Phil no respite and he still awakes, fully restored at 6am. Suicide, like the pleasures Phil desired, offers only momentary relief from his predicament. We witness Phil at his most depressed when presenting a weather prediction, "This winter is never going to end. It's going to be cold, it's going to be grey, and it's going to last you the rest of your life" (*GD*, 55:00) and indeed this is how the future must look to Phil, unable to escape his curse and unable to find happiness though the satisfaction of hedonistic desires. There appears little hope.

But Phil's many deaths do serve as a catalyst, they allow him to deconstruct himself. As Phil notes, he has been electrocuted, run over, exploded, stabbed, shot, poisoned, hanged, and burned; "I've been killed so many times I don't even exist anymore" (*GD*, 1:08:08) and from this nonexistence Phil begins to create himself anew.

§11.5 Acceptance

Phil's initial acceptance of his situation results in the claim "I'm a God" (*GD*, 1:03:02), not because he believes himself to be omnipotent (although he does appear to be immortal) but simply because he has been repeating the day for so long that he knows all of its possibilities and can manipulate and control all of the outcomes – or so he believes. Phil manages, in one 24-hour repetition, to convince Rita that something has happened to him by matter-of-factly providing intimate details about everyone in the diner and casually predicting the future. However, despite the matter-of-fact way Phil unveils his knowledge and ability there is still a degree of maliciousness in his actions; reaching the table of a young engaged couple Phil unnecessarily reveals to the groom-to-be that his fiancée is having second thoughts.

On one occasion Rita decides to spend the day with Phil, trying to understand the bizarre predicament that has beset him. However, whereas Phil sees his situation as a curse Rita is far more positive noting, "perhaps it's not a curse. It depends how you look at it." (*GD*, 1:18:00) Rita goes on to claim that she would like to have the time to repeat some things and the good that could be done through the ability to correct mistakes and prevent harm. Kupfer sees this as the moment Phil begins to change (2005, p. 291). I disagree. Phil does not appear to be convinced by Rita, making light of her claims and noting that come 6am she will have forgotten this day and will treat him with the same disdain as on every other occurrence of *Groundhog Day*. While Kupfer is correct, that from this moment Phil begins to act differently, what Kupfer seems to miss is that Phil appears somewhat detached from his actions. While outwardly Phil's actions appear virtuous it seems more likely that he is acting in this new manner simply because he has tried everything else and this, at least, provides some respite from the tedium.

I suggest the catalyst for Phil's change of character is not Rita's alternative view of his predicament. Rather, the first sign of change in Phil comes when he is forced to correct his mistaken belief that he has the power (or at least the knowledge) of a God and has the ability to control all the possible outcomes of

Groundhog Day. The discovery leads to Phil's development of a hierarchy of values, a set of values developed out of self-interest.

§11.5.1 Groundhog Day: The Ballard of the Tramp

One of the events that we see repeated time after time in *Groundhog Day* is Phil passing an individual who is down on his luck and begging passers-by for money. We are given the impression that Phil has walked past this person countless times, each time seeing him as nothing more than a nuisance. We see him patting his pockets and pretending that if he had some change he would give it but, alas, he has none. Phil shows no concern whatsoever for the tramp and on the one occasion where he does speak to the tramp it is only the ironic quip "catch you tomorrow Pops" (*GD*, 34:03)

However, after accepting his situation and following the discussion with Rita (that perhaps it is not a curse) Phil stops and hands the tramp a large sum of money. Phil does not speak to him, nor does he seem interested in the tramp, rather he appears totally uninterested. Handing over the money appears to be just something different to do and while the act appears to be one of benevolence, given Phil's knowledge that the day will repeat, the charity does not actually cost him anything. Phil's charity does not really seem to capture the essence of a benevolent act, there is no cost involved and Phil appears to lack even the remotest concern. What Phil doesn't realise is that giving the tramp money will set in motion a series of consequences and that, for the first time, Phil and the tramp will cross paths a second time in the same day. Walking home Phil sees the tramp stagger and collapse to the ground. Phil rushes him to the hospital but the tramp passes away, the nurse reports the news to Phil who refuses to accept the outcome;

[Nurse] Some people just die.

[Phil] Not today.

Phil's 'Not today' seems to express an intention to change the outcome to act so as to prevent the tramp from dying on *his* day. Rather than just giving money Phil takes the tramp in hand and feeds him, but to no avail, the tramp still dies. On another repetition Phil attempts mouth-to-mouth resuscitation, but again the tramp

dies. No matter what actions Phil puts in place he cannot change this outcome, the tramp always dies. Phil finally accepts the limitation that he cannot control others to the extent he believed.

During the many cycles where Phil vainly attempts to save the tramp, he finds the tramp becoming more and more important to him. As Kupfer notes, Phil seems to adopt the tramp, referring to him as Father, Dad, and Pop (2005, p. 291). The value Phil places upon the tramp's life grows with each repetition of *Groundhog Day*. In Randian terms Phil comes to recognise that his own life is his ultimate value and the lives of those close to him (in this case the adopted tramp) are also of great value. Phil's attempts to save the tramp are guided by self-interest. He does not want a life to be lost on 'his day', especially the life of someone who has become important to him.

In recognising that life has great value and that the cost to him (the sacrifice he must make) is in this case trivial, Phil is compelled to attempt to save the life of the tramp. There is no indication that under normal circumstances Phil will crusade to save the lives of those less fortunate than himself but things in Punxsutawney are anything but metaphysically normal¹⁰³. During the endless repetition Phil comes to value the residents of Punxsutawney, those he will interact with every day, and with this value comes a purpose. It is in Phil's self-interest to promote that which he values and, where acting in such a manner would not require the sacrifice of a greater value, Phil has reason to act benevolently towards the people of Punxsutawney.

Phil does not begin to view himself as a resource for all others as Kupfer suggests but rather recognises that life is a value and that within the confines of *Groundhog Day* he can act so as to promote that value, at little if any cost to himself. Nor, contrary to Kupfer's claims (2005, p.293), does Phil go out of his way to spare others the responsibility of being grateful to him. If anything Phil expects recognition for the sacrifices he makes as can be seen in the following case of the boy and the tree.

¹⁰³ Though we might assume Phil would be far more likely to donate a few coins.

§11.5.2 Groundhog Day: The Boy and the Tree

At some point Phil witnesses a boy fall from a tree (and presumably, as the boy falls headfirst, suffer serious injury). Recognising the value of life (within Phil's hierarchy of values) and that by way of a trivial sacrifice (just being there at a certain time) he can save the boy from substantial injury, Phil takes it upon himself to be ready to catch the boy each time he falls from the tree. But Kupfer is wrong in assuming Phil is sacrificing himself for the good of the boy and wrong in assuming that Phil does not want recognition for his act. After Phil saves the boy the child runs off without pausing to thank him, to which Phil responds:

You little brat,
You have never thanked me.
I'll see you tomorrow... maybe.

The indication is that Phil *does* expect gratitude and *does* expect recognition for his sacrifice, his taking the time (if a trivial cost) to be at the location where the boy has the accident rather than being elsewhere. Phil is not relegating himself to the position of a "resource for the benefit of others" (Kupfer, 2005, p. 293), rather he is prepared to sacrifice trivial values for greater values. That is, Phil values life just as he values other things and, on this occasion, promoting his greater value requires saving the boy.

§11.5.3 Groundhog Day: The Value of Music

Another clear illustration that Phil is acting to promote his own values and is not sacrificing himself on behalf of others is seen in his developing an appreciation of music. At some point (during the acceptance period) Phil hears music on the radio, perhaps really listening to the music for the first time, and decides to learn the piano. Beginning with no musical skill whatsoever, Phil develops into a first class pianist over the course of *Groundhog Day*.¹⁰⁴ Phil appears to take pleasure from his musical ability and musical ability is awarded a significant place within Phil's personal hierarchy of values.

¹⁰⁴ We are not explicitly told why the cycle of Groundhog Days ends and there are numerous suggested reasons such as Rita falls in love with the new Phil, or he attains true happiness. However the original script was developed by a Zen Buddhist who suggested Phil was experiencing the 10,000 cycles (some 26 years) and that the cycle of Groundhog Days had a predetermined finishing point.

According to Kupfer (2005, p. 292):

[Phil] comes to enjoy the activities for their own sake. The result is a life made meaningful by activity that is intrinsically valued. Piano playing, poetry reading, life saving – all are undertaken by Phil as ends in themselves. At the same time, Phil comes to value people for their own sake, not what they can provide him.

Certainly Phil sees the piano playing as an end in itself; the promotion of something he values simply because he values it. Further, it might be claimed that certain music is intrinsically valuable within Phil's specific worldview; the interesting aspect is the importance Phil places on this value. Each day Phil approaches the local piano teacher, who already has students booked on that day, and offers her \$1000 dollars to teach him instead. The scheduled student (a young girl) is denied her lesson, escorted by the teacher to the door, and somewhat sulkily leaves the studio. The conclusion that Phil continues to bribe the teacher and to deny the student her scheduled lesson each and every occurrence of *Groundhog Day* can be drawn, firstly because of the time it would take to develop musical skill to such a degree and secondly because the piano teacher recognises Phil as her student on the final occurrence of *Groundhog Day*, which could only occur if Phil had bribed her for lessons on that specific occurrence (even though, by this point, Phil would have little need for lessons).¹⁰⁵

In this context Phil clearly values people for what they can provide him. Phil places his value of music and musical ability sufficiently high within his own hierarchy of values that bribing the teacher and denying a young girl her piano lesson is not only justified but the recommended course of action for an ethical egoist. The young girl is denied thousands of piano lessons but only knowingly suffers the harm of being denied a single lesson. Phil clearly does not see himself as a resource for the benefit of others, does not always go out of his way to spare the feelings of others, and where his personal values are of sufficient importance is prepared to use others in order to promote these values. In terms of the axioms of ethical egoism;

¹⁰⁵ Rita, when outlining her perfect man, listed that he would play an instrument. However, Phil made no effort to learn an instrument when trying to trick Rita into a relationship and the decision to take lessons appears to be solely out of Phil's self-interest.

(i) (*Definition of Value*)

Ψ_x is a value to an agent if;

(a) Promoting or holding Ψ_x is most likely to lead to or bring about an agent's flourishing, as a certain kind of being, relative to *any* rational worldview.

or

(b) Promoting or holding Ψ_x is constitutive of what it is for an individual agent to flourish relative to that agent's worldview.

Musical ability becomes an important value for Phil to the extent that (i.b) musical ability is partly constitutive of what it is for Phil to flourish, relative to Phil's worldview.

(ii) (*Universal principle of Ethical Egoism*)

If Ψ_1 is of value to agent_n relative to agent_n's worldview

and

If Φ ing would promote agent_n's value Ψ_1

and

If Φ ing would not require the sacrifice of something in which agent_n instils greater value (Ψ_2)

then

Acting so as to promote the value Ψ_1 provides agent_n with strong moral reason to Φ , and justification for Φ ing.

Piano playing becomes of great value to Phil and advancing his musical talent promotes this value. As I have previously noted, Phil comes to value the residents of Punxsutawney, however in this case the actual harm to the girl is relatively trivial (a missed lesson) and the likelihood of Phil facing any repercussion from causing this harm is extremely minor. Phil sacrifices respect for the girl and honesty. However, given the value Phil now places on musical ability together with the fact that he cannot take a lesson on a different day (as no other days exist for Phil) not bribing the teacher would be a far greater sacrifice. Thus Phil has strong moral reason, and justification for bribing the piano teacher in order to promote the value he places in musical ability.

It may seem odd to talk of bribery as being the correct moral action but we must remember the world Phil is inhabiting is anything but metaphysically normal and I maintain that in these circumstances, given the value placed on musical ability, Phil's action *is* the moral one despite the harm done.

(iii) (*Individual motivation for acting*)

If agent_n believes that ethical egoism provides preferable reasons for acting then agent_n will, in most cases, be motivated to act in accordance with (ii).

Placing extremely high value upon his artistic outlet, Phil is motivated to bribe the teacher and deprive the girl of her piano lesson. Phil is not the sacrificial saint described by Kupfer. While on most occasions the promotion of Phil's values will result in actions that are also for the good of others (e.g. the value Phil places on life, the rescue of the boy and his attempt to save the tramp) there will be occasions where those values require the use or sacrifice of others.

§11.6 Phil and The Virtues of Flourishing-Egoism

Phil progresses from an unhappy, selfish, consequentialist egoist to a happy, fulfilled, flourishing-egoist throughout the course of one extremely long day. Along the path he develops certain virtues, those that I have claimed are necessary if an agent is to have the greatest probability of flourishing.

Phil has to adopt the virtue of honesty in order to flourish. Firstly, he must be honest with himself in accepting the bizarreness of his situation. Secondly, he is forced to be honest with himself and admit that the pursuit of immediate hedonistic pleasures doesn't bring him happiness; an observation that might be missed under normal conditions as "in ordinary life, selfish pleasures can themselves conceal what is wrong with devoting oneself to their pursuit" (Kupfer, 200, p. 293). Phil's self-honesty culminates with the realization that he does not even like himself (*GD*, 53:00), an observation that opens the path to reinventing his character. Further, Phil's real (or at least deeper) understanding of the world he inhabits only comes through being open and honest with Rita. His mistaken belief that he is an immortal God is modified and a clearer worldview developed only through honesty both with himself and with others. As Phil comes to value honesty it becomes a norm, a virtue adopted as part of his character, he is honest by default. This is not to say he is always honest, as can be seen in the example of the piano lessons; Phil has no problem in deceiving the piano teacher who believes she is giving Phil a single lesson for the \$1000 where in fact she is providing him with thousands of lessons. Honesty is sacrificed in this instance

partly because the truth would not be accepted, partly because sacrificing the truth will do no great harm (the teacher will only ever know she has given a single lesson), and partly because to be honest would mean the total sacrifice of Phil's musical ability (he will never have a single lesson, and he does not know that his Groundhog Day will at some point end). Thus, in this context, to be honest would require a greater sacrifice.

I have already described acts of limited benevolence, such as Phil attempting to save the tramp, and claimed such acts were undertaken out of self-interest. Phil's interest in the tramp comes about through the vast number of interactions he has had with him, because of his newfound values (and the recognition of life as an important value) and because Phil comes to see Groundhog Day as *his* day and he does not want someone to die on his day. We also see new acts of benevolence by Phil toward his television crew. He begins to bring them coffee and snacks prior to the early morning shoot (*GD*, 1:12:39). Partly this is done because he has adopted the virtue of limited benevolence, but again we clearly see partiality in play. Phil's success depends on the quality of his weather broadcasts and these weather broadcasts depend not just on his talent but also on the talents of his producer and cameraman. By acting benevolently towards these specific others he increases the probability that they will work to the best of their ability which, in turn, makes Phil look good. But again I stress that these acts of benevolence are not calculated or undertaken purely to gain favourable outcomes; Phil acts benevolently because it increases the probability that he will flourish, but once adopted benevolence comes naturally. However, the emotional distance between the agent and the beneficiary tempers this benevolence. Other values are likely to be given preference where the beneficiary is metaphysically distant.

During the course of *Groundhog Day* Phil also begins to value, and to take pride in, the work he does as a weatherman. We see Phil present the weather forecast many times and these reports progress from being sarcastic and uncaring to a final broadcast that is both moving and insightful. The change in the quality of Phil's weather broadcast seems to reflect the adoption of the virtue of dedication (just as it is exhibited in Phil's determination to master the piano). Phil shows dedication in taking time to research the material in his broadcast and in doing so greatly

improves the quality of the broadcast (and indirectly shows respect for both his audience and the people of Punxsutawney). Once again though, we see this respect is based in self-interest. For Phil to succeed as a weatherman requires that people enjoy his broadcasts. For people to like his broadcasts requires that Phil show respect both for the audience and the material and demonstrate that he understands the material he is presenting.¹⁰⁶ Once again, this is not a consequentialist approach; it is only through the adoption of the virtue of dedication and respect for others that Phil is able to produce the quality of show that he does and in doing so gain the appreciation of his audience – virtues are not things that can be turned on and off at whim dependent on the likely outcome of expressing those virtues.

§11.7 Conclusion

At one point in *Groundhog Day* Phil attends a session of the movie *Heidi II* dressed in the costume of Clint Eastwood's 'Man With No Name'. Kupfer (2005, p. 288) interprets this as showing the distance between Phil (like the 'Man With No Name', a friendless stranger who would like nothing more than to escape Punxsutawney) and an ideal moral character steeped in love and friendship (as represented by Heidi¹⁰⁷). How the scriptwriters or director intended this scene to be interpreted is unknown. All Rammis has to say on the reason Phil chooses to dress as the 'Man with No Name' is "because he can" (*GD*, Commentary, 40:35). However, like Kupfer, I place a great deal of significance in the scene if interpreting that significance differently. Eastwood's character was a loner and his primary motivation was always self-interest but throughout the 'Man with No Name' trilogy we see time and time again that the character has a value system in place and that there are things that he will not do regardless of how they might benefit him. I believe the symbolism expressed in this scene indicates that an individual can act from a motivation of self-interest (the Eastwood character) and at the same time embrace the virtues expressed in *Heidi* (love and friendship).

¹⁰⁶ By 'succeed' I mean to progress and perhaps gain employment with the networks, as he often indicates he desires. He is already succeeding, in as much as he works as a weatherman and his job is not under threat.

¹⁰⁷ The movie *Heidi II* is fictional and presumably a play on the sequels to the Eastwood 'Man With No Name' movies *A Fistful of Dollars*, *A Few Dollar More*, and *The Good, The Bad, and the Ugly*.

In developing a set of values, derived out of self-interest, Phil finds the reasons to respect others, to be benevolent, to be honest, to be dedicated to his chosen career, and to appreciate fine arts. Rather than sacrificing himself to others as Kupfer suggests, Phil discovers a compelling reason to be moral, namely; that acting morally is in his self-interest.

Conclusion

Out of Self-Interest

The flourishing- or value-based egoism that I have been developing is a distant cousin to classical egoism. The simple consequentialist theory, that 'what an agent ought to do on any occasion is whatever would best promote his own interests', has been developed into the following nonconsequentialist universal principle of flourishing-egoism. If Ψ_1 is of value to agent_n relative to agent_n's worldview and if Φ ing would promote agent_n's value Ψ_1 and if Φ ing would not require the sacrifice of something in which agent_n instils greater value (Ψ_2), then acting so as to promote the value Ψ_1 provides agent_n with strong moral reason to Φ , and justification for Φ ing.

This position is neo-Randian in as much as it requires an agent to act so as not to sacrifice a greater value for a lesser one. However, whereas the concept of 'value' is quite vague within Rand's work, I have analysed the notion of flourishing, commonly associated with virtue ethics, and used this as the basis for clarifying the concept of Value. Ψ_x is a value to an agent if: (a) Promoting or holding Ψ_x is most likely to lead to or bring about an agent's flourishing, as a certain kind of being, relative to *any* rational worldview, or (b) Ψ_x is constitutive of what it is for an individual agent to flourish relative to that agent's worldview.

I have claimed that for an agent to have the best chance of flourishing relative to any rational worldview the agent will necessarily value certain virtues and will adopt these virtues as part of the agent's character. In this respect a value-based egoism is elevationist in nature. Living a good life and acting with (a very limited set of) virtues is ultimately in the agent's self-interest. However, value-based egoism also has a reductionist element. The Good is reduced to an individual agent's flourishing and an individual agent's flourishing involves the promotion of that which the specific agent values.

For the most part value-based egoism promotes a course of action similar to that of the other moral theories; the egoist is compelled to save the drowning child, to share resources, and to not steal a friend's wallet. However,

egoism will at times diverge dramatically from other-regarding theories in the recommended course of action. 'Phil and the Music Teacher' (§11.5.3) provide such an example; while virtue and rights-based theories would advise that Phil ought not to bribe the teacher this is the very course of action recommended by egoism. I am uncertain how moral intuitions will run in such a case but imagine there would be considerable disagreement about what the correct course of action should be. It is where the correct course of action is uncertain that we turn to moral theories for guidance, and query 'How ought I to act?'. Flourishing-based egoism provides a clear and concise answer to this question.

I have not attempted to show that any of the other moral theories are flawed, but only to demonstrate that flourishing-based egoism is a coherent, consistent, moral theory. If the agent believes that ethical egoism provides preferable reasons for acting then the agent will, in most cases, be motivated to act in accordance with the dictates of egoism. However, where the proponents of other theories have difficulty answering the question 'Why be moral?' and often conflate the question with 'Why be moral rather than act out of self-interest?', the egoist can provide a very definite answer; you ought to be moral because being moral is in an agent's self-interest.

BIBLIOGRAPHY

Angier, N. 2004, 'No Time for Bullies: Baboons Retool Their Culture', *New York Times (Science Desk)*, 13 April 2004.

Anscombe, G.E.M. 1963, *Intention*, Blackwell, Oxford.

Baier, K. 1991, 'Egoism' in *A Companion to Ethics*, ed. P. Singer, Blackwell, Oxford.

Bayne, T.J. 2002, 'Moral Status and the Treatment of Dissociative Identity Disorder', *Journal of Medicine and Philosophy*, vol. 27, no. 1, pp. 87-105.

Blum, C.P. 1966, *Friendship, Altruism and Morality*, Routledge and Keegan Paul, Boston.

Bohse, H. 2003, 'A Paraconsistent Solution to the Problem of Moral Dilemmas', Unpublished: Presented at the 9th Annual Australasian Philosophy Postgraduate Conference, University of Tasmania.

Bok, S. 1993, 'What Basis for Morality? A Minimalist Approach', *The Monist*, vol. 76, no. 3, pp. 349-359.

Braude, S. 1996, 'Multiple Personality and Moral Responsibility', *Philosophy, Psychiatry, & Psychology*, vol. 3.1, pp. 37-54.

Brink, D.O. 1990, 'Rational Egoism Self and Others' in *Identity, Character and Morality: Essays in Moral Psychology*, ed. O. Flanagan & A. Rorty, MIT Press, Cambridge.

Brink, D.O. 1992, 'Sidgwick and the rationale for rational egoism' in *Essays on Henry Sidgwick*, ed. Bart Schultz, Cambridge University Press, Cambridge.

Broad, C. 1952, 'Certain Features in Moore's Ethical Doctrines' in *The Philosophy of G.E. Moore*, ed. P. Schilpp, Tudor, New York.

Brundle, S.E. 1996, 'Multiple Personality and Moral Responsibility', *Philosophy, Psychiatry & Psychology*, vol. 3.1, pp. 37-54.

Butler, J. 1970, *Fifteen Sermons Preached in the Rolls Chapel (1726)*, ed. J. Bernard, SPCL, London.

Carlson, G. 1973, 'Ethical Egoism Reconsidered', *American Philosophical Quarterly*, vol. 10, no. 1, pp. 25-34.

Chappell, T. 2006, 'Bernard Williams', *Stanford Encyclopaedia of Philosophy*, URL: <http://palto.stanford.edu/entries/williams-bernard/>.

Copp, D. 1997, 'The Ring of Gyges: Overridingness and the Unity of Reason', *Social Philosophy & Policy*, vol. 14, No.1, pp. 86-106

- Cordner, C. 2001, 'Ethical Necessity and Internal Reasons', *Philosophy*, vol. 76, pp. 541-560.
- Dallmayr, F. 2002, 'Asian Values and Global Human Rights', *Philosophy East & West*, vol. 52, no. 2, pp. 173-189.
- Davidson, D. 1963, 'Actions, Reasons, and Causes', *The Journal of Philosophy*, vol. 60, pp. 685-700. Reprinted in Davidson 1980.
- Davidson, D. 1980, *Essays on Actions and Events*, Clarendon Press, Oxford.
- Den Uyl, D. & Rasmussen, D. 1978, 'Nozick On the Randian Argument', *The Personalist*, vol. 59, pp. 184-205.
- Dewey, J. & Tufts, J. 1908, *Ethics*, Henry Holt, New York.
- Duska, R. 1999, 'Strictly Speaking', *Journal of Financial Service Professionals*, vol. LIII, no. 4, pp. 27-28.
- Feinberg, J. 1965, *Reason and Responsibility*, Dickenson, California.
- Foot, P. 1978, *Virtues and Vices*, Blackwell, Oxford.
- Foot, P. 1996, 'Morality as a System of Hypothetical Imperatives' in *Moral Discourse and Practice*, ed. Darwall, S. et al, University Press, Oxford.
- Frankena, W. 1987, 'Beneficence/Benevolence', *Social Philosophy & Policy*, vol. 4, iss. 2, pp. 1-21.
- Garcia, J. 1992, 'The Right and The Good', *Philosophia*, vol. 21, nos. 3-4, pp. 235-256.
- Gauthier, D. 1974, 'The Impossibility of Rational Egoism', *The Journal of Philosophy*, vol. LXXI, no. 14, pp. 439-455.
- Gauthier, D. 2006, *Morals by Agreement*, Oxford University Press, New York.
- Geach, P.Y. 1956, 'Good and Evil', *Analysis*, vol. 17, pp. 33-42
- Gert, B. 2002, 'The Definition of Morality', *Stanford Encyclopaedia of Philosophy*, URL: <http://palto.stanford.edu/entries/morality-definition/>.
- Gotthelf, A. 1999, *On Ayn Rand*, Wadsworth, United States.
- Gottlieb, P. 1996, 'Aristotle's Ethical Egoism', *Pacific Philosophical Quarterly*, vol. 77, no. 1, pp. 1-18.
- Gough, M. 1998, 'The Incoherence of Egoism', *Philosophical Papers*, vol. 27, pp. 1-28.

- Groudhog Day* (Collector's Edition) 1993, DVD, Sony Pictures, United States.
- Hadlow, N. Undated, 'Toward a Logically Consistent Kind of Ethical Egoism', URL: <http://essays/text/neilhaddow/egoism.html>.
- Hare, R.M. 1986, *The Language of Morals*, Clarendon Press, Oxford.
- Harman, G. 1977, *The Nature of Morality*, Oxford University Press, New York.
- Harman, G. 1983, 'Human Flourishing, Ethics and Liberty', *Philosophy & Public Affairs*, vol. 12, no. 4, pp. 307-322.
- Harman, G. 1986, 'Troubles with Flourishing', *Reason Papers*, no. 11, pp. 69-71.
- Hart, H. 1978, *Punishment and Responsibility*, Clarendon Press, Oxford.
- Hauptli, B. 2002, 'Kalin's "In Defense of Egoism"', URL: <http://www.fiu.edu/~hauptli/Kalin'sInDefenseofEgoism.html>.
- Hearne, V. 1987, *Adam's Task. Calling Animals By Name*, Random House, Canada.
- Heylighen, F. 1997, 'Occam's Razor', URL: <http://pespmc1.vub.ac.be/OCCAMRAZ.html>.
- Hill, T. Jr. 1999, 'Happiness and Human Flourishing in Kant's Ethics', *Social Philosophy & Policy Foundation*, vol. 16, no. 1, pp. 143-175.
- Hindman, J. 1997, 'Behind Closed Doors: [Mis]recognizing Disciplinary Gestures', URL: <http://www.utdallas.edu/pretext/PT1.1/PT1Hindmana.html>
- Hung, E. 1997, *The Nature of Science: Problems and Perspectives*, Wadsworth, California.
- Hunt, L. 1999, 'Flourishing Egoism', *Social Philosophy & Policy*, vol. 16, No. 1, pp. 72-95.
- Hurka, T. 1999, 'The Three Faces of Flourishing' in Human Flourishing, ed. E. Paul, F. Miller Jr., & J. Paul, Press Syndicate, Cambridge.
- Hurley, S.L. 1994, 'A New Take from Nozick on Newcomb's Problem and Prisoners' Dilemma', *Analysis*, vol. 54, no. 2, pp. 65-72.
- Hursthouse, R. 1999, *On Virtue Ethics*, Oxford University Press, Oxford.
- Hursthouse, R. 2003, 'Virtue Ethics', *Stanford Encyclopaedia of Philosophy*, URL: <http://plato.stanford.edu/entries/ethics-virtue>.
- Kalin, J. 1995, 'In Defense of Egoism' in *Ethical Theory*, ed. L. Pojman, Wadsworth, California.

- Kant, I. 1949, *Fundamental Principles of the Metaphysics of Ethics*, trans. Thomas Kingsmill Abbott (10th ed.), Longmans, London.
- Kant, I. 1933, *Immanuel Kant's Critiques of Pure Reason*, trans. Norman Kemp Smith (2nd ed.), Macmillan, London.
- Kavka, G. 1984, 'The Reconciliation Project', in *Morality, Reason, and Truth: New Essays on the Foundation of Ethics*, eds D. Copp & D. Zimmerman, Rowman & Allanheld, New Jersey.
- Kripke, S. 1982, *Wittgenstein on Rules and Private Language : An Elementary Exposition*, Blackwell, Oxford.
- Kübler-Ross, E. 1969, *On Death and Dying*, Touchstone, New York.
- Kupfer, J. 2005, 'Virtue and Happiness in Groundhog Day' in *The Philosophy of Film*, eds. T. Wartenberg & A. Curran, Blackwell Publishing, Oxford.
- Lear, J. 1988, 'Ethics, Mathematics, and Relativism' in *Essays on Moral Realism*, ed. G. Sayre-McCord, Cornell University Press, New York.
- Lemos, R. 1971, 'Egoism and Non-Egoism in Ethics', *The Southern Journal of Philosophy*, vol. 9, no. 4, pp. 291-392.
- Leopold, A. 2001, 'Ecocentrism: The Land Ethic' in *Environmental Ethics: Readings in Theory and Application*, ed L.P. Pojman, Wadsworth, California.
- Lycan, W.G. 1986, 'Moral Facts and Moral Knowledge', *The Southern Journal of Philosophy*, vol. XXIV, Supplement, pp.79-94.
- Machan, T. 1978, 'Was Rachels' Doctor Practising Egoism?', *Philosophia: Philosophical Quarterly of Israel*, vol.8, nos.2-3, pp. 421-424.
- Machan, T. 1979, 'Recent Work in Ethical Egoism', *American Philosophical Quarterly*, vol. 16, no. 1, pp. 1-15.
- Machan, T. 1985, 'Harman's "Refutation" of the Flourishing Ethics', *The Thomist*, vol. 49, no. 3, pp. 387-416.
- MacIntyre, A. 1984, *After Virtue*, University of Notre Dame Press, Illinois.
- MacIntyre, A. 2001, *Dependent Rational Animals: Why Human Beings Need the Virtues*, Open Court Publishing, Illinois.
- Mackie, J.L. 1977, *Ethics: Inventing Right and Wrong*, Penguin Books, London.
- Maidan, M. 1989, 'Egoism and Personal Identity', *Giornale Di Metafisica*, vol. 11, no. 2, p. 293-320.

- Matthews, S. 1998, 'Personal Identity, Multiple Personality Disorder, and Moral Personhood', *Philosophical Psychology*, vol. 11, issue 1.
- Medlin, B. 1957, 'Ultimate Principles and Ethical Egoism', *Australasian Journal of Philosophy*, vol. 35, pp. 111-118.
- Mill, J.S. 1963, 'Utilitarianism' in *Collected Works of John Stuart Mill*, eds. J.M. Robson et al., University of Toronto Press, Toronto.
- Millgram, E. 1996, 'Williams' Argument Against External Reasons', *NOUS*, vol. 30, no. 2, pp. 197-220.
- Moore, G.E. 1952, 'The Philosopher Replies' in *The Philosophy of G.E. Moore*, ed. P.Schilpp, Tudor, New York.
- Moore, G.E. 1962, *Principia Ethica*, Whitefriars Press, Tonbridge.
- Mosely, A. 2001, 'Egoism', *The Internet Encyclopaedia of Philosophy*, URL: <http://www.utm.edu/research/iep/e/egoism>.
- Nagel, T. 1970, *The Possibility of Altruism*, Princeton University Press, Princeton.
- Nagel, T. 1986, *The View from Nowhere*, Oxford University Press, Oxford.
- Nelson, M. 1999, 'Morally Serious Critics of Moral Intuitions', *RATIO, An International Journal of Analytic Philosophy*, vol. XII, no. 1, pp. 54-79.
- Nozick, R. 1971, 'On the Randian Argument', *The Personalist*, vol. 52 (Spring), pp. 282-304.
- Ostenberg, J. 1988, *Self and Others: A Study of Ethical Egoism*, Kluwer Academic, Netherlands.
- Persson, I. 1985, 'Review of Parfit, D.: Reasons and Persons', *Theoria*, vol. LI, Part 1, pp. 45-54.
- Persson, I. 1985, 'The Universal Basis of Egoism', *Theoria*, vol. LI, part 3, pp. 137-158.
- Parfit, D. 1984, *Reasons and Persons*, Clarendon Press, Oxford.
- Pietarinen, A. 2004, 'Multi-Agent Systems and Game Theory – A Peircean Manifesto', *International Journal of General Systems*, vol. 33, pp. 294-314.
- Plato, 1966, *The Republic of Plato*, trans. F. Cornford, Clarendon Press, Oxford.
- Poundstone, W. 1993, *Prisoners' Dilemma*, Anchor Books, New York.
- Putman, D. 1992, 'Egoism and Virtue', *The Journal of Value Inquiry*, vol. 26, pp. 117-124.

- Quine, W.V. 1960, *Word & Object*, M.I.T. Press, Massachusetts
- Quine, W.V. 1970, 'Grades of Theoreticity', in *Experience & Theory*, ed. L. Foster & J.W. Swanson, University of Massachusetts Press, USA.
- Quine, W.V. & Ullian, J.S. 1978, *The Web of Belief* (2nd Edition), Random House Inc., New York.
- Ramis, H. 1993, 'Audio Commentary' on *Groudhog Day* (Collector's Edition), DVD, Sony Pictures, United States.
- Rand, A. 1943, *The Fountainhead*, Bobbs-Merill, New York.
- Rand, A. 1957, *Atlas Shrugged*, Random House, New York.
- Rand, A. 1964a, 'The Objectivist Ethics' in *The Virtue of Selfishness*, ed. A. Rand, Penguin Putnam Inc., New York.
- Rand, A. 1964b, 'The Ethics of Emergencies' in *The Virtue of Selfishness*, ed. A. Rand, Penguin Putnam Inc., New York.
- Rand, A. 1964c, 'The "Conflicts" of Men's Interests' in *The Virtue of Selfishness*, ed. A. Rand, Penguin Putnam Inc., New York.
- Rachels, J. 1974, 'Two Arguments Against Ethical Egoism', *Philosophia: Philosophical Quarterly of Israel*, vol. 4, nos. 2-3, pp. 291-314.
- Rachels, J. 1978, 'Was Rachels' Doctor Practising Egoism?', *Philosophia: Philosophical Quarterly of Israel*, vol. 8, nos. 2-3, pp. 425-428.
- Rachels, J. 1986, *The Elements of Moral Philosophy* (2nd Edition), McGraw-Hill, New York.
- Rasmussen, D. 1999, 'Human Flourishing and the Appeal to Human Nature', *Social Philosophy & Policy*, vol. 16 no. 1, pp. 1-43.
- Regan, T. 2001, 'The Radical Case for Animal Rights', in *Environmental Ethics: Readings in Theory and Application*, ed L.P. Pojman, Wadsworth, California.
- Robinson, D. 2004, 'Failing to Agree or Failing to Disagree? – Personal Identity Quasi-Relativism', *Monist*, vol. 87, no. 4.
- Rogers, K. 1997, 'Beyond Self and Others', *Social Philosophy & Policy*, vol. 14, No.1, pp. 1-20.
- Rundle, B. 1993, *Facts*, Duckworth, USA.

Rottschaefer, W.A. 1999, 'Moral Learning and Moral Realism: How Empirical Psychology Illuminates Issues in Moral Ontology', *Behavior and Philosophy*, vol. 27, pp. 19-49.

Sayre-McCord, G. 1988, 'Moral Theory and Explanatory Impotence', in *Essays on Moral Realism*, ed G. Sayre-McCord, Cornell, New York.

Shaver, R. 1999, *Rational Egoism: A Selective and Critical History*, Cambridge University Press, New York.

Shaver, R. 2002, 'Egoism', *Stanford Encyclopaedia of Philosophy*, URL: <http://plato.stanford.edu/entries/egoism>.

Shubick, M. 1982, *Game Theory in the Social Sciences: Concepts and Solutions*, MIT Press, Cambridge.

Sidgwick, H. 1913, *The Methods of Ethics*, MacMillan, London.

Singer, P. 1973, 'The Triviality of the Debate over "Is-Ought" and the Definition of "Moral"', *American Philosophical Quarterly*, vol. 10, No.1, pp. 51-56.

Singer, P. 1976, 'All Animals are Equal', in *Animal Rights and Human Obligations*, ed T. Regan & P. Singer, Prentice Hall, Englewood Cliffs, New Jersey.

Singer, P. 1976, *Animal Rights and Human Obligations*, Prentice Hall, Englewood Cliffs, New Jersey.

Singer, P. 1993, *Practical Ethics*, Cambridge University Press, Cambridge.

Singer, P. 1997, *How are We to Live? Ethics in an Age of Self-interest*, Oxford University Press, New York.

Singer, P. 2001, 'A Utilitarian Defense of Animal Liberation', in *Environmental Ethics: Readings in Theory and Application*, ed. L.P. Pojman, Wadsworth, California.

Smith, M. 2003, 'Neutral and Relative Value after Moore', *Ethics*, vol. 113, pp. 176-198.

Sober, E. 1995, *Core Questions in Philosophy*, Prentice Hall, New Jersey.

Solomon, R.C. 1995, 'Nihilism', in *The Oxford Companion to Philosophy*, ed T. Honderich, Oxford University Press, New York.

Stingl, M. 2000, 'All the Monkey's Aren't in the Zoo: Evolutionary Ethics and the Possibility of Moral Knowledge', *Canadian Journal of Philosophy*, Supplementary vol. 26, pp. 245-266.

- Stroud, S. 1998, 'Moral Overridingness and Moral Theory', *Pacific Philosophical Quarterly*, vol. 79, pp. 170-189.
- Sturgeon, N. 1988, 'Moral Explanations', in *Essays on Moral Realism*, ed G. Sayre-McCord, Cornell University Press, New York.
- Thomas, L. 1993, 'Moral Flourishing in an Unjust World', *Journal of Moral Education*, vol. 22, Iss. 2, pp. 83-98.
- Tucker, A. & Kuhn, H. (eds.) 1950, *Contributions to the Theory of Games*, Princeton University, New Jersey.
- Unger, P. 1996, *Living High and Letting Die: Our Illusion of Innocence*, Oxford University Press: Oxford.
- Van Ingen, J. 1994, *Why Be Moral?*, P. Lang, New York.
- Walker, M.U. 2000, 'Naturalizing, Normativity, and Using What "We" Know in Ethics', *The Southern Journal of Philosophy*, vol. XXVI, Supplement, pp.75-101.
- Wilkes, K.V. 1988, *Real People: Personal Identity without Thought Experiments*, Clarendon Press, Oxford.
- Williams, B. 1981, 'Persons, Character, and Morality' in *Moral Luck*, Cambridge University Press, Cambridge.
- Wittgenstein, L. 2001, *Philosophical Investigations*, trans. G.E.M. Anscombe (3rd ed.), Blackwell, Oxford.
- Wolf, U. 1986, 'Was es heisst, sein Leben zu leben', *Philosophische Rundschau*, vol. 33, pp. 247-274.
- Zemach, E. 1978, 'Love Thy Neighbor as Thyself or Egoism and Altruism', *Midwest Studies in Philosophy*, vol. 3, pp. 148-158.