

**The Metaethical Ramifications of Empirical Moral Psychology:
Examining the Moral Cognitivist/Non-Cognitivist Debate**

by

Jonathan Courtney

A thesis submitted to the Faculty of Graduate and Postdoctoral
Affairs in partial fulfillment of the requirements

for the degree of

Master of Arts

in

Philosophy

Carleton University

Ottawa, Ontario

© 2012

Jonathan Courtney



Library and Archives
Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

ISBN: 978-0-494-93631-3

Our file Notre référence

ISBN: 978-0-494-93631-3

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

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

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Abstract

In this work, I look to explore the metaethical ramifications of moral psychology by examining the impact data from moral psychology has on the cognitivist non-cognitivist debate. I found that the data from moral psychology suggests that both cognitive and non-cognitive mental states are involved in moral judgments, and this in turn puts pressure on more extreme cognitivist and non-cognitivist positions to adopt a more moderate position in order to remain empirically plausible.

Acknowledgements

I would like to acknowledge the numerous individuals who helped me in the production of this work. First and foremost I would like to thank Professor Gordon Davis for his role in supervising this work. I would also like to thank Professor Heidi Maibom and Professor David Matheson for inspiring many of the ideas that found their way into this work. I would also like to thank my fellow students, and in particular, Iain Laidley and Mary Renaud for their helpful discussions and support.

I would also like to acknowledge the financial contributions that made this thesis possible. These contributions came from the Social Sciences and Humanities Research Council via the Joseph-Armand Bombardier Canada Graduate Scholarships Program Master's Scholarships, The Ontario Government via the Ontario Graduate Scholarship Program, and Carleton University via a number of Departmental Scholarships.

Table of Contents

ABSTRACT	II
ACKNOWLEDGEMENTS	III
TABLE OF CONTENTS	IV
INTRODUCTION.....	1
CHAPTER 1 UNDERSTANDING MORALITY- WORKING BETWEEN METHODOLOGIES.....	6
1.1 THE CASE OF EXPERIMENTS AND PHILOSOPHY.....	7
1.2 WORRIES ABOUT EXPERIMENTAL PHILOSOPHY	8
1.3 IN DEFENCE OF EXPERIMENTAL PHILOSOPHY.....	13
1.4 THE IMPLICATIONS OF THIS DEBATE.....	17
CHAPTER 2 EXAMINING THE MORAL COGNITIVISM/NON-COGNITIVISM DEBATE	18
2.1 MORAL COGNITIVISM AND MORAL NON-COGNITIVISM.....	18
2.2 MOTIVATIONS FOR MORAL COGNITIVISM AND MORAL NON-COGNITIVISM	20
2.3 ALLAN GIBBARD’S NON-COGNITIVISM.....	23
2.4 MICHAEL SMITH’S COGNITIVISM	28
2.5 COGNITIVISM/NON-COGNITIVISM- TWO CENTRAL THESES.....	32
CHAPTER 3- CONNECTING MORAL PSYCHOLOGY AND THE COGNITIVIST NON-COGNITIVIST DEBATE.....	36
3.1 WHAT IS IT FOR A MENTAL STATE TO BE COGNITIVE/NON-COGNITIVE?	36
3.2 COGNITIVISM AND NON-COGNITIVISM IN THEORIES OF THE EMOTIONS	37
3.3 COGNITIVE VS NON-COGNITIVE MENTAL STATES: A SET OF CRITERIA	39
3.4 DOES THAT MENTAL STATE <i>INFLUENCE</i> OR <i>CONSTITUTE</i> THAT MORAL JUDGMENT?	46
3.5 CONSULTING THE DATA: HAIDT, GREENE AND MOLL	47
CHAPTER 4- JONATHAN HAIDT- THE SOCIAL INTUITIONIST MODEL.....	49
4.1 EXPOSITION.....	49
4.2 EVIDENCE IN FAVOUR OF THE VIEW.....	55
4.3 CRITICISM: SALTZTIEN AND KASACHOFF	67
4.4 RESPONSE TO SALTZSTEIN & KASACHOFF.....	72
4.5 SIM’S IMPLICATIONS FOR PSYCHOLOGICAL NON-COGNITIVISM	78
CHAPTER 5- GREENE’S DUAL-PROCESS MODEL	84
5.1 EXPOSITION.....	84
5.2 EVIDENCE	86
5.3 CRITICISM: MCGUIRE ET AL.....	91
5.4 RESPONSE TO MCGUIRE ET AL.	93
5.5 IMPLICATIONS FOR PSYCHOLOGICAL NON-COGNITIVISM.....	95
CHAPTER 6 MOLL- THE EVENT-FEATURE-EMOTION-COMPLEX VIEW	101
6.1 EXPOSITION.....	101
6.2 EVIDENCE	104

6.3 CRITICISM: CASBEER AND HYNES	106
6.4 RESPONSE TO CASBEER AND HYNES	109
6.5 IMPLICATIONS FOR PSYCHOLOGICAL NON-COGNITIVISM	110
6.6 TO WHAT EXTENT DO HAIDT GREENE AND MOLL SUPPORT PSYCHOLOGICAL NON-COGNITIVISM?	118
CHAPTER 7- THE IMPACT OF MORAL PSYCHOLOGY ON THE MORAL COGNITIVIST/NON-COGNITIVIST	
DEBATE	123
7.1 THE EFFECTS OF PNC* ON MODERATE POSITIONS.....	124
7.2 DOUBLE EXTERNALISM AND ITS OPPOSITION TO PNC*	127
7.3 OBJECTIONS TO PNC*	130
7.4 GENERAL IMPLICATIONS OF PNC*	136
CONCLUSION.....	137
APPENDIX A- ADDITIONAL CONSIDERATIONS FOR HAIDT AND GREENE.....	141
Notes.....	144
Bibliography	154

Introduction

What should we do? In the context of day to day conversation, this simple question is used in order to get advice on some course of action toward some particular end. Just behind the simple colloquial use of this phrase lays a profound philosophical problem. How do we determine what we ought to do? In other words, what ends or values should direct our actions and govern our behaviour? In this work, I will do little to address these grand and foundational questions. Instead I will focus on some of the problems that arise downstream from the detailed exploration of these questions. In particular I will look to compare and partially reconnect two sub-disciplines that have grown out of the investigation of these questions. In particular I will attempt to show how the findings of moral psychology can have an important impact on particular questions in the domain of metaethics. In order to explain the relevance of this thesis, it will be important to take a few steps back from this claim and look at the methods that surround the investigation of moral questions in contemporary philosophy.

As those familiar to contemporary debates in moral philosophy will be well aware, the study of morality has become a large and complex enterprise. The complexity and diversity of moral questions have split our discussion of morality into several sub-disciplines. Traditionally, moral philosophy is divided into three sub-disciplines, metaethics, normative ethics and applied ethics. Applied ethics focuses on analyzing particular cases of moral judgment. This field looks to analyze issues like abortion, animal rights, capital punishment and environmental rights, and determine to what extent certain kinds of action or policy are morally permissible. While these particular controversial topics are often at the forefront of popular consciousness, applied ethicists

generally rely on any number of more general and less widely known ethical systems in order to reach their conclusion. Normative ethics investigates these broader ethical theories and attempts to justify why one way of systematizing moral claims is better than another. These debates generally focus on particular moral theories such as utilitarianism and deontology, and attempt to demonstrate that one system is better able to capture, explain and systematize our moral intuitions. Metaethics is one step further detached from particular moral issues, and it looks to analyze the meaning of our moral concepts themselves. Common metaethical debates include questions concerning the metaphysical, epistemological and semantic considerations surrounding moral concepts. Particular debates include debates surrounding whether or not moral truth exists, whether or not moral knowledge is possible, and whether or not moral discourse should be considered 'truth sensitive' in the first place. Together these three fields represent the traditional focuses philosophers have taken when they investigate moral claims. While each of these fields has a different focus, they each look to understand morality using the traditional methods of philosophy such as introspection and conceptual analysis. In other words, they attempt to answer questions about morality by analyzing the constraints of our moral concepts and intuitions, and trying to work out implications for the nature of morality from these constraints.

In addition to these three more traditional fields of philosophical investigation, there is a new field which attempts to investigate morality from another perspective. Empirical moral psychology looks to understand questions of morality in terms of empirically testable claims about human psychology. Instead of trying to deduce features about morality from our intuitions or features of our moral concepts, empirical moral

psychology attempts to understand morality as a psychological phenomenon which can be understood using the methods of empirical investigation. This methodology attempts to answer questions about what morality is, how it operates and how it motivates us by appealing to data from behavioural psychology, social psychology and neuroscience. This approach is novel because it allows us to take a perspective on morality which does not rely on the traditional philosophical method of conceptual analysis. That being said, empirical moral psychology is an explicitly descriptive project, and so it will not be concerned with certain features of the study of morality. It focuses on specific descriptive claims about how morality originates in the human subject and how it affects our behaviour, and it explicitly ignores questions concerning the justifiability of our moral claims and the aptness of certain moral concepts.

Each of the four sub-disciplines I have addressed so far attempt to carve out a particular area of moral investigation. It is a commonly held belief among ethicists that each of these sub-disciplines operates roughly independently from one another. I argue though that this conventional notion fails to appreciate the extent to which conclusions reached in one domain have the potential to affect the others. While I think clear links can be drawn between applied ethics, normative ethics and metaethics, for the purposes of this project I will not focus on these connections, but on a connection between two positions which are, in many ways, on the opposite ends of the spectrum in their methodological commitments. The connection I will attempt to demonstrate in this project is the connection between metaethics and empirical moral psychology. It is relatively rare in the contemporary literature for philosophers to draw explicit connections between these two methods for investigating morality. This may have to do

with the fact that groups have traditionally pursued fundamentally different methodologies. Metaethics has been concerned with complex and abstract conceptual analysis, while empirical moral psychology is concerned with testable observable empirical facts about human behaviour. Despite the seeming distance between these two methods, I argue that there are several debates in which these methods could overlap. In particular, contemporary metaethicists commit themselves to several positions which I believe can be translated into substantial psychological claims which in turn can be tested by the methods of empirical moral psychology. In this work, I will develop this approach by looking at one debate in particular which draws this connection. By analyzing this debate, I will look to address many of the methodological and conceptual difficulties involved in moving from one domain of moral inquiry into the other. That being said, I am confident that, insofar as metaethicists and empirical moral psychologists can be shown to be talking about the same phenomenon, they can each provide substantial insights into each other's fields of inquiry.

In this project, I will be focusing on the debate between the moral cognitivists and moral non-cognitivists. Moral cognitivism is roughly the view that moral judgments are claims about the world.¹ Moral Non-Cognitivists on the other hand hold that moral judgments are not concerned with claims about the world, but are concerned primarily with the expression of some feeling or non-belief state. Over the course of this work I will attempt to demonstrate that the data from moral psychology can have an important

¹ taken in the broadest sense, to include claims about relations between actions and mental responses, and also to include claims about normative properties that some take to stand apart from the natural world.

impact on this debate. In the first chapter I will consider some concerns surrounding the use of empirical data in philosophical inquiry and I will look to clearly define the context in which this data can affect philosophical conclusions. Once I have accomplished this task, I will move on to my next chapter, where I will consider the cognitivist and non-cognitivist debate in more detail. In this chapter I will use considerations from the first chapter to identify precisely which features of the cognitivist/non-cognitivist debate are amenable to being assessed in the light of the empirical data put forward by moral psychology. Once I have clearly highlighted this feature of the debate, I will move on to Chapter 3, where I address potential concerns over how we can apply the data from moral psychology to the relevant feature of the cognitivist/non-cognitivist debate. In Chapters 4, 5 and 6, I will consider the accounts of three contemporary moral psychologists, Jonathan Haidt, Joshua Greene, and Jorge Moll. These three individuals' findings will provide a good overview of much of the data available in the contemporary debate. By comparing the conclusions of these three accounts, I will be able to draw a general conclusion about how data from contemporary moral psychology affects the cognitivist non-cognitivist debate. This conclusion, roughly, will be that moral judgments are constituted by both cognitive and non-cognitive mental states. In Chapter 7 I will analyze the implications of this conclusion for the contemporary debate, focusing in particular on the pressure it places on more extreme positions in the debate.

Chapter 1 Understanding Morality- Working between Methodologies

In this work, I will be looking to evaluate the extent to which the moral cognitivist non-cognitivist debate can be affected by data from moral psychology. In the introduction, I briefly discussed what I take to be the methodology and focus of both metaethics and moral psychology. Before we can begin applying the findings of empirical moral psychology to the cognitivist non-cognitivist debate, we must first address a substantial concern that looms in the background of this investigation. This concern has to do with the extent to which empirical data can be applied to questions in philosophy. In order to address this concern I will focus on a debate where these controversies are currently at play. To this end I will focus on the debate surrounding the controversial and growing field of *experimental philosophy*. By analyzing this debate, we will be able to get a concrete look at many of the controversies surrounding the use of empirical data on philosophical problems. In an attempt to capture this debate, I will be focusing on a set of objections provided by Antti Kauppinen. As we will see, Kauppinen's objections will focus on the idea that there is a set of questions that empirical investigation cannot properly address. These questions have to do with the constraints of our concepts, and experimental philosophy's primary error is its attempt to investigate these questions empirically. Nichols and Knobe respond to this objection and argue that despite Kauppinen's conclusion, empirical considerations still have an important role to play in our philosophical investigations. Considerations from this debate will lead us to some general conclusions concerning the role of empirical data in philosophical inquiry. These considerations will be helpful in framing the rest of our investigation.

1.1 The Case of Experiments and Philosophy

Experimental philosophy is a field that attempts to reach conclusions about philosophical concepts using the tools of experimentation rather than those of introspection or conceptual analysis. While philosophy has traditionally attempted to determine the intuitive appeal of particular philosophical positions through introspective analysis and debate, experimental philosophy adopts an empirical method that looks to determine the extent to which the public in general accepts or rejects some feature of a particular philosophical concept. Since experimental philosophy is primarily concerned with understanding what every-day individuals' intuitions are on particular philosophical issues, they generally gain their empirical data through surveys which are generally completed either by undergraduate philosophers, or anonymous online test subjects. While numerous interesting insights have come out of experimental philosophy it has not been without its critics. In order to address some of these potential concerns about this methodology, I will focus on the critique of experimental philosophy offered by Antti Kauppinen in "The Rise and Fall of Experimental Philosophy".¹ After considering each of Kauppinen's central arguments, I will evaluate their plausibility in the light of some arguments in favour of experimental philosophy as presented by prominent experimental philosopher's Shaun Nichols and Joshua Knobe. Considerations from this debate will in turn provide us with some general conclusions about the use of empirical data in philosophical debates.

1.2 Worries about Experimental Philosophy

In his paper “The Rise and Fall of Experimental Philosophy” Antti Kauppinen forwards a number of compelling objections against the project of experimental philosophy.² Kauppinen’s central claim is that the methodology of experimental philosophy is flawed because it is committed to the fallacious assumption that the results of a survey can properly test the viability of our intuitions. As he puts it “Experimentalists interpret claims about intuitions as straightforwardly empirical and therefore testable predictions about how ordinary people will answer when presented with actual or hypothetical cases.”³ Kauppinen argues that intuitions, and philosophical concepts properly construed, are not ‘straightforwardly empirical’ and as such, a methodology based on this assumption cannot be adequate. In order to understand the basis of Kauppinen’s critique, it will be necessary to discuss in more detail what he takes the experimental philosophical method to be. Kauppinen provides following formalized version of this method:

(E) “In S, we would (not) say that X is C’ is a prediction that (most) non-specialists will (not) say that X is C if the case S is presented to them.”⁴

Here S refers to a particular scenario or thought experiment, X refers to a particular element of that scenario or thought experiment and C refers to a particular concept which may or may not be properly applied to X. Thus the experimental philosopher looks to understand the proper application of C by examining how most people will use this concept when confronted with a particular scenario. While this may seem like a reasonable method at first, Kauppinen argues that it is problematic because it relies on the faulty assumption that we can reach conclusions about the proper application of our

concepts simply by appealing to how most individuals use these concepts. This assumption is problematic because our philosophical concepts are *elliptical*, meaning that in order to understand them properly we need to be familiar with certain background considerations. His claim is that while in an academic philosophical context individuals have the abilities and background knowledge necessary to make proper judgments about philosophical concepts, individuals outside this context lack these important conditions. Kauppinen therefore claims that simply polling individual's general intuitive reactions to thought experiments fails to accomplish the central task of (E), which is to understand the proper constraints of the concept being investigated. As a way of focusing this criticism, Kauppinen, proposes the following proper methodology for discussing our intuitions about a concept:

(A) 'In S, we would say that X is C' is a hypothesis about how (1) competent users of the concepts in question would respond if (2) they considered the case in sufficiently ideal conditions and (3) their answer was influenced only by semantic considerations.⁵

Kauppinen lists these three conditions in (A) to ensure that the intuition of the agent in question tracks the semantic constraints of the concept being investigated. Each of the three conditions of (A) are designed to exclude certain short comings that might prevent normal subjects from tracking the semantic constraints of their concepts. (1) Echoes our earlier concern about agent's inability to fully grasp the constraints of the concepts they are employing. While most individuals will likely qualify as competent users for more simple concepts, complex philosophical concepts might only be partially understood by many members of the public. Thus when we are considering whether or not a concept

applies in a particular case, we should limit ourselves to the opinions of those individuals who fully understand the concept. Criteria (2) and (3) of (A) echo concerns we might have about the conditions individuals are in when they answer the surveys used by experimental philosophers. For example, we would want individuals to have the opportunity to consider the application of the concept in question free of the distractions or time constraints that may be imposed by a survey. Furthermore, we would not want individuals to be distracted by pragmatic linguistic practices at play in the question being asked. For example, an individual might interpret the claim ‘the individual went to lunch’ as a claim about the agent doing so freely, as the pragmatic linguistic practice is to simply assume that an agent is acting of their own volition unless otherwise specified. This pragmatic linguistic practice could interfere with questions concerning intuitions about freedom or other discussions about the nature of agency. Thus (2) and (3) both look to insure that the agent’s analysis of the application of a concept stays true to the concept itself and is not distorted by any features of the survey process.

Since there is no guarantee that the three ideal features articulated in (A) will obtain in (E), Kauppinen argues that the experimental approach can provide no real insights into the nature of our shared philosophical concepts. Kauppinen argues that in order to insure the existence of the three conditions present in (A), the experimental philosopher would have to take measures to insure that the subject understood the concept and were applying that concept correctly. Such measures though would require the philosopher to engage with the individual in a manner that is excluded by the method of an impartial survey. Kauppinen argues that the best method for such interactions would be a sort of Socratic Method, which would ask the interlocutor questions to insure

that they are using their concept consistently, using proper counterfactual analysis, and focusing on the semantic and not the pragmatic features of the question at hand. This conclusion would strongly oppose the methods of experimental philosophy, as the focus of the analysis would no longer be on what people think, but instead on what people properly primed and directed would think. This priming is designed to replicate the sort of reflective conceptual analysis that normally takes place in traditional forms of moral philosophy. As a consequence, actually going through the Socratic process with an individual is unnecessary and we can simply revert back to the forms of analysis that traditionally comprise philosophical discourse. For these reasons Kauppinen concludes that experimental philosophy is a failed disciple, as it could only reach a substantive philosophical conclusion if it followed the constraints of (A), and in doing so it would simply revert back to the traditional forms of philosophical investigation.

As we have seen, Kauppinen's objections are focused on the methods and conclusions of experimental philosophy, and so one might fairly pose the question, what does all of this have to do with moral psychology? As we saw in the introduction, empirical moral psychology is a field that attempts to address questions about morality using data from empirical psychological investigation. While many of Kauppinen's objections were focused on problems particular to experimental philosophy, many of the general arguments Kauppinen provided for rejecting (E) could affect the methods of moral psychology as well. Recall that one of Kauppinen's objections to (E) is that it fails to appreciate the fact that we need a competent judge in order to properly delineate where a concept applies. This failure could potentially affect the methods of moral psychology, as many experiments often require subjects to distinguish between moral and non-moral

cases. If standard individuals are not competent judges, then the data collected on 'moral' phenomenon might be tracking some other psychological feature, as the subjects might be considering features that have nothing to do with a proper understanding of morality. A similar worry could arise from concerns about ambiguities in the questions individuals are asked in these experiments. These ambiguities could get in the way of testing the subject's ability to differentiate between moral and non-moral cases. In addition to these procedural concerns, Kauppinen's critique raises a larger and potentially more troubling objection to the methods of empirical moral psychology. This more significant worry can be derived from the broader criticism that motivates much of Kauppinen's discussion. This criticism concerns the role of empirical data in matters of conceptual analysis. It holds that certain considerations can only be determined by the traditional philosophical methods of introspection and conceptual analysis. These areas can only be investigated using these methods because they are concerned primarily with how we structure our ideas, and not with how these ideas relate to particular facts about the world. This more general objection criticises experimental philosophy on the grounds that it confuses a field of inquiry that is concerned with structuring our ideas with one which looks to represent certain facts about the world. While most fields of inquiry are concerned with both aspects of conceptual analysis and matters of empirical fact, it is reasonable to worry that we might mistakenly assume that certain empirical facts can speak to certain conceptual distinctions, or that certain conceptual distinctions necessitate certain empirical facts. Thus as we continue this analysis it will be important to insure that conclusions being drawn from moral psychology do not commit us to this error. Before addressing this broader concern, it will be helpful to consider some possible

replies to Kauppinen's objection to experimental philosophy. As we will see, these responses will sharpen Kauppinen's more general concern about the role of empirical data in philosophical domains. After focusing in on this concern I will attempt to elucidate those conditions in which empirical data can address the difficult philosophical problems pursued by experimental philosophy.

1.3 In Defence of Experimental Philosophy

An individual who is sympathetic to the project of experimental philosophy could adopt a number of different strategies in order to respond to Kauppinen's position. The first would be to respond to his points directly by arguing that his portrayal of experimental philosophy is too pessimistic, and that features 1-3 of (A) could be accommodated in an adequately structured survey. While there is certainly ground to argue that Kauppinen is rather unfairly pessimistic about the ability of everyday people to understand philosophical concepts and represent them accurately on a survey, I believe that there is a more basic and more decisive strategy for responding to Kauppinen's objections. This strategy would argue that Kauppinen's position is based on an improper representation of the goals of experimental philosophy. According to this objection, the formulation of the experimentalist's approach which Kauppinen provides in (E) is far narrower than the position defended by contemporary experimental philosophers. As we have seen in the last section, Kauppinen's main criticism of the experimental philosopher is that they expect truths about the nature of our concepts to be determined by what most people think about these concepts. Putting the point crudely, Kauppinen accuses the experimental philosopher of doing conceptual analysis by appeal to popular opinion.

Despite Kauppinen's seeming claims to contrary, I argue that most experimental philosophers would also be opposed to this method of understanding our concepts. The modus operandi of experimental philosophy is not to infer truths about our concepts from what most individuals think about the concept. Instead, experimental philosophers are interesting in determining how most individuals think about philosophical concepts in order to gain access to a distinct, though intrinsically related type of knowledge.

Experimental philosophers are looking to understand not only the conceptual constraints of our concepts, but the psychological processes that lead us to adopt our concepts in the first place. As Knobe and Nichols put it, experimental philosophers are methodological pluralists, and believe that both the method of experimentation and the method of conceptual analysis have a role to play in the proper understanding of philosophy.⁶

Once we have identified that the experimental philosopher is interested in a different set of questions than traditional conceptual analysis, it becomes clear that most of the objections raised by Kauppinen no longer apply. For instance, it is no longer a concern that most individuals aren't competent users of a concept, as explaining the psychological systems responsible for the systematic misuse of a concept is as interesting from an empirical perspective as understanding those systems behind the proper application of the concept. Indeed, many of the most interesting studies in moral psychology demonstrate that most individuals' moral judgments can be influenced significantly by aspects of a scenario which seem to be entirely missing from our conscious moral concepts. Such scenarios may tell us nothing about the necessary and sufficient conditions of the moral concepts of the individuals in the experiment, but they none the less tell us a great deal about what actually impacts individuals moral behaviour

or moral judgment. Given these considerations, it is clear that the experimental philosopher can agree with Kauppinen's assertion that (E) is a poor way of understanding our concepts, while also preserving the project of experimental philosophy as one which looks to understand those psychological systems which produce these concepts.

While this response helps to resolve the primary objection forwarded by Kauppinen, it also opens up the experimental philosopher to a new set of objections. These objections focus on criticising the new method being used by the experimental philosopher. As we have seen, this method focuses on understanding the psychological processes that produce our philosophical concepts, and then using this information to come to new conclusions about the concepts themselves. Nichols and Knobe consider a number of objections to this new method in their article "An Experimental Philosophy Manifesto."⁷ The first and perhaps most important objection they consider is that this new method cannot yield any interesting philosophical data. This objection holds that while this method might reveal some interesting psychological facts about how we cognize certain philosophical concepts, it is incapable of providing us with any useful information about the concepts themselves. The reason it fails to tell us anything about the concept itself is the fact that the content of a concept is not necessarily tied to the psychological system that produces it. Thus even if the experimental philosopher can reach definitive conclusions about the psychological systems that produce our philosophical concepts, they will fail to say anything about the concepts themselves. Since this method doesn't tell us anything about the philosophical concept it is purporting to investigate, the experimental philosopher cannot provide information relevant to philosophical problems.

While certain features of this argument are appealing, it is based on an overly rigid distinction between those problems that can be addressed by conceptual analysis and those that can be addressed by empirical investigation. In particular, the objection fails to appreciate that many of our philosophical positions are directly related to claims about the psychological systems that produce them. For example, one way of interpreting debates about the plausibility of our intuitions is to view them as debates over whether the intuitions represent particular features of the world accurately. According to this perspective, we trust our intuitions because we believe they accurately represent some feature of the world. Thus it is clearly important to this debate to determine whether or not the psychological systems that produce our intuitions do so consistently and accurately, or whether they are subject to mass deception and error. As Nichols and Knobe put it: “Surely the degree to which an intuition is warranted depends in part on the process that generated it, and surely the best way to figure out which processes generate which intuitions is to go out and gather empirical data. How else is one supposed to proceed?”⁸ Thus we see that facts in the domain of psychology are relevant to philosophical investigation whenever our philosophical positions commit us to claims about human psychology. Philosophical positions can commit themselves to truths about human psychology indirectly, as in cases where we assess the aptness of the psychological systems that produce our philosophical intuitions, or more directly, as in cases where philosophical positions make direct claims about the nature of the psychological systems associated with our philosophical concepts. For example, much of contemporary philosophy of mind is concerned with theories about language acquisition, complex cognition and consciousness. Each of these areas of inquiry make significant

claims about the nature of certain shared psychological systems, and as such the plausibility of these theories become tied to what we learn about how these systems function. Thus, experimental philosophy can clearly exonerate itself from the charge of being concerned with data which is ‘philosophically uninteresting’ as many of our philosophical positions are tied to facts about the psychological systems that produce them. This in turn provides a strong defence of the method of experimental philosophy, as it can not only resist the objection articulated by Kauppinen, but it can also provide a new method for discovering philosophically interesting conclusions.

1.4 The Implications of this Debate

In considering the debate surrounding experimental philosophy we have seen that there is much controversy surrounding the use of empirical data in philosophical pursuits. That said, our defence of the experimental philosopher’s position has brought us to a more general conclusion about the nature of empirical investigation in philosophical affairs. In particular, this analysis has led us to the conclusion that empirical investigation is relevant in any field of philosophy which commits itself, either implicitly or explicitly, to certain claims about the world. These considerations put us in a good position to begin our discussion about how the empirical data from moral psychology can affect the cognitivist non-cognitivist debate. In particular it directs our attention toward attempting to demarcate a feature of this debate which is committed to some empirically testable claim about human psychology. In the next chapter, we will turn our attention toward the moral cognitivist/non-cognitivist debate, and attempt to articulate which features of this debate could be affected by the findings of empirical moral psychology.

Chapter 2 Examining the Moral Cognitivism/Non-Cognitivism Debate

We ended the last section with the conclusion that empirical moral psychology can in principle address philosophical positions when those positions are committed to particular claims about human psychology. In this section we look more closely at the debate surrounding moral cognitivism and moral non-cognitivism, and attempt to determine whether or not the debate commits itself to any particular claims about human psychology. I will begin this task by building on the brief description of the debate provided in the introduction. After presenting the two positions in more detail, I will provide some context to help explain the motives that drive each position. I will then move on to provide a contemporary example of each position, Allan Gibbard's non-cognitivism and Michael Smith's cognitivism. After appreciating the complexities of these contemporary positions, I will go on to consider which features of this debate are amenable to the observations of empirical moral psychology.

2.1 Moral Cognitivism and Moral Non-Cognitivism

Moral cognitivism is the position that moral judgments are beliefs. A belief, for our purposes, can be roughly defined as an attitude that affirms a particular proposition. In the philosophical context 'propositions' occupy a number of different roles including being the referents of that-clauses, providing the meaning of sentences and being the bearers of truth value.⁹ Thus, according to the moral cognitivist, moral judgments have to do with making a claim about the world that has a particular structure, and which is truth-sensitive. While this distinction may seem obtuse or abstract, it really only delineates a

very simple feature of our language which is at play in many of our everyday concepts. For example, when I make the claim “water is wet” I am making a particular claim about the world that could be true or false. In this case, I am predicating the property of 'wetness' onto water. Moral cognitivists want to argue that a similar mental process is occurring when we make our moral judgments. The cognitivist would argue that when I make the judgment 'murder is wrong' I am attributing the property of 'wrongness' to the act of murder. Just as the claim about water being wet commits me to the world being a particular way, the claim that murder is wrong also commits me to the world being a particular way. Each of these claims has a truth condition, and as such each of these claims is shown to be a belief.

Moral non-cognitivism, much as the name would entail, is the denial of this position. Non-cognitivists hold that moral judgments are not beliefs and as such have no truth conditions. Instead the non-cognitivist attempts to explain moral judgments by conceiving of them as non-truth apt speech acts such as an expression of an emotion. One of the earliest forms of non-cognitivism was emotivism. Emotivism is the position that moral judgments are not claims about the world, but a particular kind of emotional response we have to particular behaviour. For example, the emotivist would interpret the claim 'murder is wrong' as 'Boo!: murder' wherein the 'Boo!' represented the expression of a negative emotion. These sorts of claims are not apt for truth or falsity because they do not make any claims about the world. While most contemporary non-cognitivists don't give a story as crude as the emotivist regarding the interpretation of moral language, they all agree that what is at the heart of a moral judgment is not a propositional description of the world but some sort of expression or non-propositional speech act.

2.2 Motivations for Moral Cognitivism and Moral Non-Cognitivism

In considering the history of the debate between the moral cognitivist and the moral non-cognitivist, it is helpful to consider that moral cognitivism has largely been the default position in the debate. This should not be particularly surprising as the majority of our moral discourse appears to commit us to claims about the world that seem to go beyond mere expressions of our feelings or preferences. Richard Joyce makes the point in the following manner: "...imagine a member of a hospital ethics committee expressing her judgments as a series of "Hurray's" and grunts of disapproval- we would be appalled." Indeed we would treat such behaviour "not merely as inappropriate but as utterly mystifying."¹⁰ Given these observations, we might wonder what has led philosophers to adopt such a seemingly problematic position.² Both in the history of the position, and in more contemporary cases, individuals defend moral non-cognitivism in response to certain problems that seem to be entailed by thinking of moral properties as features of the world. Emotivism was one of the first versions of non-cognitivism and it was introduced as a way of explaining morality within the methodological constraints of logical positivism. Roughly, logical positivism is the position that all knowledge falls into one of two categories, the tautological claims of mathematics and logic, or empirically verifiable claims of science. For the logical positivists, any claim that did not fit into one of these two categories was considered nonsensical. Those logical positivists

2 It should be noted that contemporary examples of non-cognitivism are much more sophisticated than early emotivism, as will be seen in discussions of Gibbard's positions. In addition it should be noted that contemporary accounts that look to put the emotions at the center of our moral judgments do a great deal to address the objections articulated by Joyce and others.

who addressed questions of morality were motivated by problems like Moore's Open Question Argument³ to conclude that moral properties could not be natural properties. Since moral properties could not be the natural features of the world discoverable by science, and since they were not mathematical or logical, all talk of morality must be seen as empty and guilty of some fundamental conceptual error. While some logical positivists committed themselves to an Error Theory⁴ concerning moral judgments, others were not content in simply labeling morality a nonsensical 'empty set'. These individuals wanted to provide some explanation for why morality seemed to be meaningful to us, and why it played the role it did in human affairs. In order to provide such an analysis, logical positivists such as A.J. Ayer¹¹ and C. L. Stevenson¹² advanced the theory of emotivism. Emotivism attempted to solve the problem by appealing to a feature of human behaviour that was not governed by the conceptual constraints imposed by logical positivism. Since emotive expression was non-propositional, it did not make any claim that attempted to be true, and as a consequence did not have to fit into either the category of logical tautology or empirically verifiable fact. One can emotively express 'Boo!: Murder' without attributing any property to the act of murder, and since there is no property, there is no

³ Moore's Open Question Argument, is an argument advanced by G.E. Moore in *Principia Ethica*, which purports to show that no natural property could function as the definition of 'the good'. Roughly, this argument is based on the observation that, for any natural property X, the question 'is it true that X is the good?' is open, which is to say that it is not clearly the case. Moore argues that if X were the proper definition of 'the good' the question, would be 'closed' such that it would be obvious that the two were equivalent. Since the question of X's equivalency to 'good' is open, Moore, and many philosophers to follow him, concluded that 'the good' could not be a natural property.

⁴ Error Theory is a theory which holds both the semantic thesis that all moral claims refer to moral truth makers and the ontological thesis that no moral truth makers exist. Thus the Error Theorist treats moral discourse in a manner similar to the way atheists would interpret statements about God. Such discussions intend to refer to something, but no such thing exists.

concern about failing to meet the rigorous constraints of logical positivism. Conceiving of moral claims as non-propositional, and thus non-cognitive, was thus first adopted by the logical positivists in order to help navigate the difficulties of an extremely strict ontology in conjunction with scepticism concerning the plausibility of moral naturalism.

While the above discussion explains why the logical positivists were motivated to adopt non-cognitivism, we are left to wonder why those of us who are not logical positivists should accept this position. While logical positivism isn't considered a very plausible position in contemporary debates, many of the considerations that drove the logical positivist to adopt non-cognitivism still serve to motivate contemporary non-cognitivist accounts. In particular, contemporary non-cognitivists share both the logical positivists' commitment to properties being empirically detectable, and their skepticism concerning natural facts' ability to act as propositional content for moral judgments. Contemporary non-cognitivists are motivated by problems like Moore's Open Question argument, and Hume's is-ought problem to attempt to provide an alternative explanation of the content and normative force of our moral judgments. These individuals look to provide an explanation of our moral judgments qua non-propositional in order to address these concerns. Unlike the positions that came before them, contemporary non-cognitivists are sensitive to the fact that our moral language appears to be cognitive and propositional in some sense. These contemporary accounts attempt to provide an alternative explanation of our seemingly cognitive moral judgments. In doing so they hope to demonstrate that the key features of our moral judgments are more akin to an expression of an attitude, as opposed to a propositional claim about the world. If they are successful in this endeavour, the non-cognitivist will explain how moral judgments can

be non-propositional, and hence avoid the problems of Moore and Hume, while still being firmly committed to a modest form of naturalism. In the next section, we will examine how the non-cognitivist might provide such an explanation of moral judgment by considering Allan Gibbard's highly influential metaethical account.

2.3 Allan Gibbard's Non-Cognitivism

Allan Gibbard provides one of the most famous and theoretically robust contemporary accounts of non-cognitivism. Gibbard has provided several instantiations of this account in his career, and in this section I will be examining this position as it evolved from his earlier work, *Wise Choices Apt Feelings*¹³ into his most recent adaptation in *Reconciling our Aims*.¹⁴ In his latest work, Gibbard advances the position that normative questions are essentially concerned with actions and plans for actions. He suggests "that we think of ought questions, in the first instance, as planning questions."¹⁵ But what does it mean for questions of normativity to be about planning? The idea is roughly that at the most basic level, when we talk about what we *ought* to do, we are talking about plans for action. The assertion of a particular plan is then seen to be the key feature that makes a particular discourse normative. While all normative judgments have to do with plans, moral judgments in particular are plans that interact with our moral sentiments. So for Gibbard "what is it to think an act morally wrong, as opposed to just silly or imprudent ... is to think that the act warrants resentment on the part of others and guilt on the part of the person who did it."¹⁶ Gibbard here aligns himself with the sentimentalist tradition, which posits that moral judgments are constituted, at least in part,

by particular emotional responses. Without such emotions responses these positions claim, we would not be making genuine moral judgments.

We have seen that for Gibbard, moral judgments are assertions of a particular course of action that are associated in the right way with our moral sentiments. It is important to note how this is consistent with the general constraints of non-cognitivism. The preference of a particular plan can be reinterpreted as the statement 'let's do x!' This statement is not a proposition, and as such is not capable of being made true or false by the world. Instead it is simply an assertion of an agent's preference of one course of action over another. In addition, our moral sentiments do not commit us to any propositional claims about the world, as they are simply expressions of a particular feeling in relation to a particular event. Thus, when we interpret moral claims in this manner, we see that a claim like 'murder is wrong' is simply an expression of two different attitudes: the expression 'Don't murder!' and, a negative emotional response ('yuck!: murder' 'boo!: murder'). Gibbard believes that these two features can explain the content and normative force of our everyday moral discourse.

While this line of explanation of our moral judgments seems consistent enough, it still needs to explain why it seems, on the face of things, that we are making propositional judgments when we make moral claims. For example, when I make the claim 'x is wrong' I appear to be attributing the property of 'wrongness' to 'x'. In addition it seems to make sense to talk about moral judgments using truth-sensitive language. For example, when I make the claim 'x is wrong' in a moral context, I am not merely intending to express my opinion, I am intending to make the claim that *it is true* that 'x is wrong.' In order for Gibbard's position to be plausible, it has to account for these

seemingly truth-sensitive features of our moral discourse. One strategy Gibbard could adopt to try and address these features of moral language would be to first show how certain plans commit us to particular factual claims about the world, and then demonstrate how these factual claims can account for the propositional and truth-sensitive component of our discourse. This alternative semantic account can start with the observation that inconsistent plans ('Do X!' and 'Don't do X!') cannot both be accepted as a courses of action. When we discuss two different courses of action we open ourselves up to truth-sensitive language, because these plans in part commit us to certain beliefs about the world. If a plan commits me to being both in Ottawa and on top of Mt. Kenya at the same time then the plan cannot be accepted, as it conflicts with the rather plausible belief that I cannot be in two places at the same time. So there is a sense in which any moral claim can be false if it commits us to a course of action which is impossible or which is conditional on false claims about the world. Since our plans are concerned with the world there is a sense in which they are truth sensitive, as each of our plans are premised on the world being one way rather than another. While there is, in a sense this referential component to our plans, it is important to note that the normative or moral force of these plans does not arise from this referential component, but from our desire to have a functional "complete contingency plan of living."¹⁷ Thus, while truth sensitive-language may work its way into our moral discourse, it is derivative of our non-truth-apt expressed preference of a set of plans.¹⁸

While this may constitute one strategy whereby Gibbard could make room for truth-sensitive language in our moral discourse, it is clear that there is a great deal of moral discourse that appears to be truth sensitive that does not commit us to any non-

moral facts about the world. Mundane examples such as ‘murder is wrong’ appear to be truth sensitive and involve predication, without relating to some factually problematic claim embedded in a set of plans. In order to provide a more robust response to this problem, we will have to try another approach. In order to provide this approach it will be necessary to focus on some of the details Gibbard provides in an earlier instantiation of his account as presented in *Wise Choices, Apt Feelings*.

In this work, Gibbard introduces his explanation of moral discourse using a thought experiment involving the Goddess Hera who is “entirely coherent and completely opinionated both normatively and factually.”¹⁹ According to Gibbard, Hera would have both a fully determinate understanding of the way the world is (w) and a coherent system of norms which she accepts (n). Together these two sets would constitute what Gibbard calls a “factual-normative world w,n.”²⁰ In this “factual-normative world” there is a truth, not only about the way the world is, but also about what is to be done in every situation. Describing things in this manner allows us to introduce a new form of predication which Gibbard calls ‘n-predication’. N-predication arises from our acceptance of a particular set of norms, and can be applied to claims that are normatively relevant. When considering a particular normative statement (S) we can re-interpret the normative commitments embedded in S to be n-predicates of a particular normative-factual world. Blackburn provides the following description of Gibbard’s translation of our moral claims:

“Now to settle whether normative statement S holds in factual-normative world w, n we do the following. Replace each normative predicate in S with its n-corresponding descriptive predicate. That yields a purely

descriptive statement S. Then normative statement S holds in w, n if and only if S_n holds in w."²¹

Since this analysis reduces the normative components of S to a series of descriptive claims about a set of norms, we can then evaluate the truth of S in the context of a particular factual-normative world. The claim, 'A ought to do x' holds in w,n if and only if 'A is n-required to do x' is true in w. We can therefore interpret the discussion of 'truth' and 'falseness' in moral claims as claims about what would or would not obtain in a particular normative-factual world. In addition we can explain the seemingly predicative nature of moral language by appealing to n-predication. According to this view, if I make the claim 'it is true that we ought to give more money to charity' I am, in fact, expressing my endorsement of a normative-factual world n,w where 'we are n-required to give more money to charity' is true. By making the focus of normative discourse normative-factual worlds, Gibbard provides a non-cognitivist way of explaining the seemingly truth-referential and propositional features of moral language. The reason this explanation remains a non-cognitivist one is that the features being predicated by moral language are not features of the world, but features of a set of plans that the agent accepts. The foundation of normative discourse then is not a truth about the world, but an expression of a set of values held by an agent, encapsulated in a general plan for action. While we may represent this plan propositionally, it is our expressed preference of this plan, not a fact about the world, that makes our moral claims what they are.

Now that I have given a rough presentation of Gibbard's non-cognitivist account of moral judgment, I will move on to review an equally well-regarded and insightful philosopher of the opposing position. Michael Smith provides an excellent example of a

contemporary cognitivist view which addresses many of the same concerns that Gibbard attempted to address in his account. Like Gibbard, Smith provides a detailed and insightful account of the nature of moral judgment, and also like Gibbard he adopts certain features that seem to at first favour his opponents' position. As we will see, despite the fact that both of these individuals take opposing views on the question of the nature of moral judgment, their accounts have many important similarities. It will be important to clearly articulate these similarities and differences in order to enable our future investigation concerning the psychologically-testable claims being made by these two positions.

2.4 Michael Smith's Cognitivism

In *The Moral Problem* Michael Smith provides a defense of his version of moral cognitivism in the context of addressing what he takes to be the problem at the heart of contemporary metaethical debates. Smith argues that many of the difficulties and complexities of the current debate originate from the fact that moral claims have to exhibit two seemingly mutually exclusive behaviours. First, they have to behave like beliefs about the world, things which seem to be true or false. Second, they need to have practical implications, and should *ceteris paribus* motivate us to act in one way rather than the other. The conjunction of these features is problematic because it conflicts with a relatively plausible and widely accepted position in traditional moral psychology. According to this view of psychology, originally articulated by Hume, there are two distinct kinds of psychological states: beliefs and desires. Beliefs purport to represent the way the world is, while desires are states that represent how we want it to be.²² In

addition, according to this view motivation only arises when we couple a particular desire with a particular means-end belief. The problem that emerges for our explanation of moral judgments then is that they seem to need to possess *both* features of a desire and a belief. The conjunction these features is explicitly denied by the Humean Theory of Motivation. Given Smith's general acceptance of the Humean Theory, the problem that emerges is to try and explain how these three mutually exclusive commitments can operate together. Smith formalizes these premises as:

1. "Moral judgments of the form 'It is right that I ϕ ' express a subject's beliefs about an objective matter of fact, a fact about what it is right for her to do.
2. If someone judges that it is right that she ϕ 's then, *ceteris paribus* she is motivated to ϕ
3. An agent is motivated to act in a certain way just in case she has an appropriate desire and a means-end belief, where belief and desire are, in Hume's terms, distinct existences."²³

Smith argues that much of debate in contemporary metaethical discourse comes down to an attempt to address these three mutually exclusive premises. One way of resolving the tension between these three positions would be to refute the Humean Theory of Motivation. If we could allow for a mental state that could act as both the source of motivation and a factual representation of the world, what Smith calls a 'besire', then we could reject 3, and still accept 1 and 2. Smith considers at length possible responses to the Humean Theory of Motivation, and for the sake of this paper it is only necessary to note that Smith provides considerable objections to each of these responses. Thus,

according to Smith, if we are going to solve the Moral Problem, we are going to have to focus on the first or second premise in the problem.

Another way to solve the problem would be to go the non-cognitivist route and simply deny the first premise by arguing that moral claims are not beliefs. Smith provides a number of reasons why he thinks that non-cognitivism cannot capture the main features of our moral discourse. For the purposes of this work it won't be necessary to investigate the details of any of these objections. It will be sufficient to note that Smith's commitment to cognitivism arises from some substantive arguments he presents against the non-cognitivist. The last possibility then is to reject premise two. Rejecting premise two would commit one to moral judgment externalism, the position that one can sincerely hold a moral judgment in normal conditions and not feel any motivation pull toward a particular course of action. Smith again provides a number of arguments to show why he takes this position to be implausible.

After defending these three premises, the only option left to Smith is to show how these three seemingly mutually exclusive premises can operate together. In order to accomplish this task, Smith sets out to show how a particular set of beliefs can provide us with reasons and motivation for action. Smith begins his exploration of this possibility with the observation that there is an important difference between *desiring* something and *valuing* it. There is a gap for example between what the kleptomaniac 'resolves' to do after a process of deliberation (not steal and hoard items) and what he wants to do (steal and hoard items). This gap demonstrates that an agent can desire something without valuing it, which is to say that we can have a desire to act in a particular way without believing we have reasons to act in that manner. From this information, Smith concludes

that our values are a particular kind of belief. These beliefs are normative because they are necessarily connected with our desires. The reason Smith links our moral beliefs to our desires is because under the Humean Model, only desires can provide an agent with motivation for action. Smith goes on to conclude that there is one belief in particular that bears this special relationship with our desires. This is the belief “that what we have normative reason to do is what we would desire that we do if we were fully rational.”²⁴

According to this position, what makes a claim like 'x is wrong' true, is that I would not do x in that situation if I were fully rational. In Smith's words “To say that we have a normative reason to ϕ in certain circumstance C is to say that we would want ourselves to ϕ in C if we were fully rational.”²⁵ How though does this premise of full rationality bridge the gap between the factual features of beliefs and the motivating features of desires? The idea is that while a desire is always involved in the motivation for action, in the case of a normative judgment it is not enough for us to merely have a desire; we want to be justified in having this desire. If we found out, for example, that our desire was based on a false belief, or conflicted with a more general desire that we took to be more important, we would want our desires to change. Smith formalizes this platitude of our everyday judgments, and argues that what we really want is not simply what we desire, but what we would desire if we were fully rational. This, in a sense, bridges the Humean gap between desires and beliefs by providing us with a belief 'that if we were fully rational we would ϕ ' which leads to a desire. It is important to note that this is not calling into question Hume's division between beliefs and desires, it is merely presenting us with a particular kind of belief which pushes all rational agents toward a particular set of desires. By defining a normative reason as 'what we would want ourselves to do if we

were fully rational' Smith also allows for a cognitivist reading of our moral judgments. We appear to be making truth-apt propositional claims when we engage in moral discourse because we are in fact predicating a particular set of properties; namely that set of things which we would desire if we were fully rational.

Smith's position is an interesting one to consider for the sake of this project, as it is one that accepts the importance of desire in our everyday motivation, while also favouring a cognitivist reading of moral judgments. Smith would not be opposed to desire, or other non-cognitive features being involved in the motivations surrounding our moral claims. While moral judgments are ultimately beliefs for Smith, their motivational force comes entirely from their ability to interact with our desires. The key point then is that while much of what is happening *around* moral judgments is strictly non-cognitive, the moral judgments themselves should be interpreted in a cognitive manner. Given that cognitivists about moral judgment can allow for a great deal of non-cognitive psychological activity to surround moral judgments, it will be a particularly demanding task to separate out precisely what the distinction is between non-cognitive elements influencing moral judgments, and non-cognitive elements constituting moral judgments. In particular, it will be difficult to articulate precisely what, from an empirical psychological perspective, makes these views distinct.

2.5 Cognitivism/Non-Cognitivism- Two Central Theses

As we have seen, Gibbard and Smith each provide a robust and complex explanation of the nature of our moral judgments. The complexity of these accounts has emerged as a result of the long and fruitful debate that has taken place between these two

views on the nature of moral judgment. It will be important to keep in mind the complexities of the two accounts, and to appreciate that a clear distinction between the contemporary cognitivist and contemporary non-cognitivist position is muddled by the fact that each position has adopted certain features of its opponents' view. In particular, contemporary non-cognitivist accounts generally include some form of pseudo-predication in their moral accounts, while many cognitivists such as Smith tie the propositional content of their moral views quite closely to the non-propositional components of human psychology. These features of contemporary accounts will make it increasingly difficult to keep the two separate, particularly in the context of testing their empirically verifiable claims. In order to move forward in this analysis then, it will be necessary to attempt to articulate some clear distinction that can separate these two positions. In particular, we will need to identify those features of the debate that are amendable to empirical investigation. In order to accomplish this task I will rely on the analysis of these two positions provided by Mark van Roojen in his article "Moral Cognitivism vs. Moral Non-Cognitivism".²⁶

In his discussion of the debate between the moral cognitivists and the moral non-cognitivists, van Roojen focuses on two commitments made by the non-cognitivist that help to separate the two positions. The first is *semantic non-factualism*, which denies that moral sentences, properly conceived, express propositions or have truth conditions. This feature of the non-cognitivist position is akin to the emotivists' assertion that moral judgments are not so much statements, as they are expressions of approbation or disapprobation. As we have seen, contemporary accounts like the one articulated by Gibbard attempt to provide some form of pseudo-predication to explain our use of moral

terms. That said, even these accounts would support *semantic non-factualism* by arguing that no genuine truth-referential predication is occurring when we make moral judgments. *Semantic non-factualism* then is largely a claim about the proper definition of moral judgments, and whether the way we use moral terms requires us to think of them as a form of predication.

As we observed in the last chapter, claims concerning the proper constraints of our language are best addressed by standard methods of conceptual analysis. As such if *semantic non-factualism* was the only thesis that divided the cognitivist from the non-cognitivist, then there would be little hope in investigating this distinction using data from moral psychology. While this data might provide some insight about the mechanisms that produce our semantic content, such an analysis would miss the object of concern, which is the semantic content itself. Fortunately for my project, there is another distinct, though closely related position that divides the cognitivist from the non-cognitivists. This additional position is the latter's commitment to *psychological non-cognitivism*. *Psychological non-cognitivism* is an explicitly psychological thesis which "denies that the states of mind conventionally expressed by moral utterances are beliefs or mental states which fall on the cognitive side of the cognitive/non-cognitive divide."²⁷

As we concluded in the previous chapter, philosophical positions become amenable to the insights of empirical moral psychology when they commit themselves to particular claims about human psychology. *Psychological non-cognitivism* clearly commits itself to a particular claim about the nature of the mental states involved in moral judgments, and as such is amenable to the findings of contemporary moral psychology. That being said, it is not entirely clear what it would mean for a moral

judgment to “fall on the cognitive side of the cognitive/non-cognitive divide.”²⁸ As such, in order to continue this analysis we need a clear understanding of what this distinction amounts to in the domain of psychology. Once we have become clearer on this distinction further work will need to be done in order to assess whether data from contemporary moral psychology can speak clearly to this distinction. In the next chapter, I will focus on addressing these concerns in a manner that will enable the rest of this investigation.

Chapter 3- Connecting Moral Psychology and the Cognitivist Non-Cognitivist debate

In the last chapter, we concluded with the observation that the moral cognitivist and the moral non-cognitivist are each committed to particular claims about human psychology. In rejecting *psychological non-cognitivism* moral cognitivists commit themselves to the claim that the mental states constituting moral judgments are ‘cognitive in the way beliefs are’. In defending *psychological non-cognitivism*, the non-cognitivists commit themselves to the opposite psychological hypothesis, that mental states are not ‘cognitive in the way beliefs are’ but expressions of non-propositional attitudes. While *psychological non-cognitivism* does make a particular claim about the nature of human psychology, it is not entirely clear from a psychological perspective what this claim is. Thus, before we can consult this data effectively, we need to have a clearer understanding of what it would mean in the domain of psychology for a judgment to be ‘cognitive in the way beliefs are’. The first section of this chapter will therefore be dedicated to establishing what, from a psychological perspective, the distinction between a cognitive and a non-cognitive mental state amounts to. As we will see, our conclusions about this distinction will have far reaching consequences for the rest of our investigation.

3.1 What is it for a mental state to be cognitive/non-cognitive?

The first major ambiguity in the question we are investigating surrounds a precise psychological definition for what it is for a mental state to be cognitive rather than non-cognitive. This distinction in the metaethical literature is important because it influences the ontological and semantic commitments of our moral discourse. If it turns out that our

moral judgments predicate properties then we need to determine what these properties are and to what extent they exist in the world. When we shift our focus from the ontological and semantic distinction to its psychological counterpart, it becomes less clear precisely what the division entails. In order to clarify how the distinction between cognitivism and non-cognitivism might translate into the domain of moral psychology, it will be helpful to look at how the division is presented in a similar debate. The debate I am referring to is the debate surrounding the mental processes involved in the generation of the emotions. In this debate, one of the main distinctions is between the cognitive and the non-cognitive theories of emotional processing. While these positions are not directly related to the positions we have been discussing thus far, they bear sufficient similarity to serve as a helpful entry point into the psychological debate.

3.2 Cognitivism and Non-Cognitivism in Theories of the Emotions

Cognitive theories of emotional judgment hold that emotional responses are intrinsically linked to certain cognitive mental processes.²⁹ These theorists argue that our emotions are always constituted, in part, by a certain conscious judgment or appraisal of a situation. This position is motivated by the observation that our emotional reactions are often a product of the beliefs or judgments we have about the situation we are reacting to. Two individuals facing the loss of their job might have entirely different emotional responses as a result of their perspective on being unemployed. Likewise, two dramatically different situations might evoke the same emotional response insofar as they lead two individuals to adopt the same belief about how their situation will affect them. These observations lead the emotional cognitivist to assert that some other cognitive

process, like the activation of some cognitive judgment, must take place before any emotional response is possible.

Non-cognitive theorists conversely argue that judgments or appraisals are not a primary feature of an emotional response. They argue that an emotional response follows directly from the perception of a relevant scenario and does not require any additional judgment or appraisal.³⁰ This position can be thought of as attempting to represent the folk psychological notion that emotions are irrational, and largely uncontrollable. They are thought of as ‘hot’ and automatic, while more cognitive processes are considered cool and directed. Non-cognitive theories look to defend this conception of the emotions by arguing that higher-order cognitive activity is not necessary for our emotional responses. Different non-cognitive theorists rely on different mental processes to help explain the production of emotions without the use of reason or judgment. Paul Ekman for example posits two systems to explain the generation of our emotions, the *appraisal mechanism* and the *affect programme*.³¹ The *appraisal mechanism* operates quickly and automatically to identify the emotionally relevant features of the scenario. When the relevant stimuli are present the *appraisal mechanism* triggers the appropriate *affect programme*. The *affect programme* governs the muscular and neurophysiological changes associated with the particular emotion. These automatic processes then function without the agent’s knowledge or control to elicit the emotion in question.³² This simplistic portrayal of Ekman’s position is sufficient to communicate the central thesis common to non-cognitive theories of emotion. This central thesis is that emotional responses do not depend on higher-order cognition or judgment, and instead arise from psychological processes that are automatic and largely out of the agent’s control.

The debate between these two positions is a substantial one. That said, for our purposes it will be necessary to look past many of the substantive points of contention and focus instead on the debate's categorization of *cognitive* and *non-cognitive* mental states. Using observations from this debate, I will now attempt to isolate a few key features that will help us to delineate between cognitive and non-cognitive mental states. In doing so I will focus on outlining those features of cognitive and non-cognitive mental states which are both relevant to the moral cognitivist/non-cognitivist debate and which fit the general distinction present in the debate surrounding the emotions.

3.3 Cognitive vs Non-Cognitive Mental States: A set of Criteria

As we have seen, the cognitive theorist in the domain of the emotions focuses on features of our psychologies that are conscious, and have to do with beliefs and judgments that are largely propositional in form. The cognitive theories posit that the judgments and appraisals that trigger our emotions are the sort of mental entities that occupy our daily deliberations, the sort of things that we consciously represent to ourselves as reasons for action. These mental entities are not only transmittable through language, but are also for the most part concerned with accurately representing certain features of the world. Given these observations, the first main feature we can attribute to a cognitive mental state is that it is *concerned with propositions*. These states are concerned with propositions because they are concerned with those features of our thought that can be systematically structured and fully communicated through language. For example, when I state 'the cat is on the mat' I am communicating a particular piece of information concerning the cat being on the mat. Upon hearing this phrase, the listener

has gained full access to the proposition contained within my belief. In contrast if I say “Boo!: that cat on the mat” in a particular tone, I cannot directly communicate the emotional valance of the 'Boo' feature of this sentence.⁵

While this distinction is difficult to clearly articulate in psychological terms, it will be an important distinction to highlight for our purposes, as it is one of the primary features which separates cognitive and non-cognitive theories of moral judgment. Part of the difficulty in articulating this distinction from a psychological perspective comes from the fact that it is not often discussed in the domain of moral psychology. As a consequence an important amount of interpretive work will need to be done to divide the psychological data that exists into the categories of ‘propositional’ and ‘non-propositional’ mental states. This interpretive work will be one of the main goals of the chapters ahead.

The observation that cognitive mental states are concerned with propositions leads us to the second feature of cognitive mental states, namely that *they involve beliefs*. As we observed in the last chapter, beliefs are propositional attitudes that take the proposition they contain to be true. As we observed, cognitivists about emotion hold that a proposition like ‘there is a lion right behind me’ only serve to elicit an emotion if we in some sense take that proposition to be true. If for instance, an individual only wondered whether or not a lion was behind them, it would only invoke the emotion of fear in proportion to the perceived likelihood of the lion being there. This same characteristic can be observed in the moral cognitivist portrayal of moral judgments. Once an individual makes a moral judgment, they are not simply considering a moral proposition,

⁵ although the individual may empathize with my sentiment and also 'Boo' at the cat.

but take it to be true. Thus a cognitive mental state is one that is not only propositional in structure, but also endorsed by the individual who adopts it. Thus we can conclude the second criteria of a moral judgment, which is that *it involves a belief*.

I argue that the two features we have discussed thus far are the primary features of cognitive mental states. I will focus on these two psychological features in particular, because they align with both the commitments of the cognitivist in the domain of the emotions and the general features of moral cognitivism. While these two features will be the main focus of the rest of this analysis, there are some other psychological features that divide cognitive from non-cognitive mental states. These secondary features arise from the ways in which the mind processes propositional versus non-propositional mental states. The first of these features is the fact that cognitive mental states generally operate slower than their automatic non-propositional counterparts. When we are consciously working through a particular problem, we are interacting with propositional beliefs we have about the subject at hand. Even basic forms of this kind of conscious rational thinking take a long amount of time, particularly in relation to other more automatic brain functions. It takes longer for an individual to work their way through a math problem for the first time than for an individual to understand a phrase spoken in their native tongue. One way of explaining the difference between these two facts is the fact that the former is a conscious deliberative process while the latter is an automatic unconscious process. While it is certainly up to debate to what extent fast-acting automatic processes can also be propositionally structured, cases of conscious rational deliberation are prime candidates for propositionally structured mental states.

These considerations lead to an additional secondary feature of cognitive mental states, that they are consciously accessible and in normal conditions controllable by the agent. Judgments and beliefs are, in general, the kind of mental state which we take ourselves to be able to access and control. While there may be some beliefs that are not consciously accessible to us, the vast majority of our beliefs are accessible and, to some extent, in our control. If a system of information were both totally inaccessible to us and completely beyond any conscious control, it would likely not be a very good candidate for a propositional belief. Thus a good candidate for a secondary feature of cognitive mental states is that they are consciously accessible and capable of being controlled. There are of course some potential concerns for this secondary feature. As we will observe later in the discussion, there are several psychological theories which posit that propositional beliefs can exist both outside of our awareness and control. While it is beyond the scope of this thesis to make a substantive claim about the plausibility of these theories, it is important to note that even if they are plausible, it does not take away from the fact that propositional mental states are the primary candidates for consciously accessible and controllable mental states. Non-propositional mental states are not, for the most part, the sorts of things we can deliberate and reason about in the traditional sense. Thus when a mental state is one which we actively deliberate over in a rational manner, the mental state is most likely to be propositional. Just as in the case of the speed of the mental process, the accessibility and controllability of the process only imperfectly correlates to the primary features we are concerned with. As such, these positions are both best construed as secondary features. These observations can lead us to a general list of the features a mental state will satisfy if it is cognitive:

<p><u>Primary</u> 1. Propositional 2. Involve Beliefs</p>	<p><u>Secondary</u> 3. Relatively slow 4. Consciously accessible and controllable.</p>
--	---

Now that we have constructed a rough list of the features a cognitive mental state will possess, we can move on to consider what different features we should expect to characterize non-cognitive mental states. Observations from Ekman's position can lead us to our first conclusion, that non-cognitive mental processes operate quickly and without the subject's conscious awareness. As we saw, Ekman's *affect programme* operated quickly and automatically outside the agent's awareness and control. Thus, non-cognitive mental states can be seen to have the inverse of cognitive mental states secondary features; they are both fast acting and occur largely without the subject's awareness. Like the secondary features of the cognitive mental states, the secondary features of non-cognitive mental states are derived from these states' primary features. The first primary feature of these states is that they are non-propositional. This feature is primary because it captures the central distinction that is relevant to the moral cognitivist non-cognitivist debate. The reason we can know that non-cognitive mental states are non-propositional is because they function to process information in a manner that is entirely distinct from the agent's consideration of propositions. Like Ekman's *affect programme*, these mental states are designed to respond to certain stimuli and evoke certain structured responses without entering into the agent's set of beliefs. Thus these mental states can be thought of as non-propositional because they process information in a way that is not belief-like, and does not lend itself to being expressed in propositional language. As a consequence, non-cognitive mental states are not beliefs, in the sense that they do not contain

propositional content that is concerned with truth or falsity. Thus we can see from this short discussion that these processes have the inverse features of the cognitive mental states. That is to say non-cognitive mental states can be characterized by the following features:

Primary: 1. Non-propositional 2. Not concerned with beliefs	Secondary: 3. Operate quickly and automatically, 4. Not consciously accessible or controllable
--	---

These considerations provide us with a rough set of criteria for distinguishing cognitive from non-cognitive belief states. As we have seen, these general psychological desiderata parallel nicely with the features that divide moral cognitivists from moral non-cognitivists. Now that we have a general set of criteria, we are better equipped to address contemporary data from moral psychology, and determine whether or not it speaks in favour or against *psychological non-cognitivism*. That said, before we can carry out this analysis, we need to address a few further methodological concerns. The first of which has to do with a potential objection to the list of secondary features proposed for cognitive and non-cognitive mental states.

One argument that could be raised against the discussion so far is that my divide between cognitive and non-cognitive mental states enforces an artificial divide in the nature of human psychology. One way of articulating this concern would be to point out that despite what I have said to the contrary, there are cases where automatic non-conscious psychological processes can be propositional. Certain propositional attitudes could become so ingrained through habit that they manifest themselves automatically in

the relevant circumstances. Such attitudes could be cognitive in the sense that their content is truth-apt, and have a distinct communicable linguistic meaning, but they may operate automatically and without conscious deliberation. For example, certain mathematical truths such as the fact that $2+2=4$, can become so habituated that we automatically reach the conclusion without conscious deliberation. Thus automatic mental processes need not be non-propositional.

This is a reasonable concern, and highlights an important point, namely that my division of psychological states between 'cognitive' and 'non-cognitive' is not meant to cover all possible psychological processes. There are certainly cases of propositional beliefs becoming habituated, and thereby becoming automatic, but this does not undermine my distinction. As I mentioned before, I list these as secondary qualities precisely because they are imprecise predictors for the division I am concerned with. I am concerned first and foremost with dividing the two sets of mental states along the lines of propositions. I argue that many non-propositional states are automatic, precisely because they do not enter into the agent's set of beliefs and as a consequence aren't the sorts of things to be consciously considered. In addition, I have defined cognitive mental states as conscious and deliberative because propositional claims monopolize consciously accessible deliberative reasoning. While there are arguably many cases of propositional reasoning that happen automatically, most kinds of belief states are easier to identify in conscious and deliberative psychological activity.

The discussion of this objection will be important to keep in mind in the upcoming analysis of the data. As we will see, much of the contemporary literature in moral psychology points to automatic psychological processes playing an important role

in moral judgments. In light of this objection it will be important to not equivocate 'automatic' with 'non-cognitive' as we have seen that propositional claims can also be cognized automatically. Thus in the final analysis, the primary focus will have to be on whether the automatic processes being discussed by contemporary moral psychology are propositional or non-propositional in nature.

3.4 Does that mental state *influence* or *constitute* that moral judgment?

Before we can move on to assess the data provided by contemporary moral psychology, it will be necessary to become a bit clearer on what *psychological non-cognitivism* is committed to. In particular, the point that needs to be stressed is that *psychological non-cognitivism* is a thesis about the nature of the mental states that *constitute* moral judgments, not those that merely *influence* them. This distinction is important because it changes the nature of the data that is necessary to affirm *psychological non-cognitivism*. For example, we could not vindicate this thesis merely by demonstrating that non-cognitive affective features are involved in moral judgments, as these features could be involved in our judgments, even if our judgments are purely propositional. An individual may feel disgust toward a scenario, and then form a conscious rational belief on the basis of this disgust. In this case the non-cognitive mental state was important only insofar as it triggered the cognitive judgment that followed it. The moral judgment in this example was constituted by a cognitive mental state and so while the process of forming the judgment implicated a non-cognitive mental state, the example would not count in favour of *psychological non-cognitivism*. In the metaethical literature many cognitivists are comfortable with non-cognitive mental states

playing a role in moral judgment. As we saw, Smith was closely tied to Hume's observations about moral psychology, and as such was amendable to the idea that the sentiments play a key role in motivating moral judgments. None the less, Smith would argue that the content of moral judgments themselves are cognitive, and in particular are constituted by the belief 'that which we would do if we were fully rational.'

Moving forward then, we should narrow our analysis to the following constraints. First we must focus on identifying the mental states that constitute our moral judgments. Once these mental states have been isolated we should test whether or not they have the four descriptors associated with cognitive or non-cognitive mental states, focusing particularly on whether or not the mental states are best categorised as propositional or non-propositional in nature. Now that we have a general method of investigation, we can go on to consider the best way in which to address the significant amount of data presented by empirical moral psychology.

3.5 Consulting the Data: Haidt, Greene and Moll

Our discussion so far has provided us with a general framework for deciding what data we should focus on in the moral psychological literature. Unfortunately, even with a relatively clear set of criteria for our analysis it is difficult to address the significant amount of moral psychological data that is available. In anticipation of this difficulty I will focus my analysis on the positions advanced by three particular moral psychologists: Jonathan Haidt, Joshua Greene and Jorge Moll. By focusing on these individuals positions I will be able to rely on their abilities to synthesize a large set of empirical data into a particular theory. Using these theories, we can determine whether or not the data

they present favours *psychological non-cognitivism*. I have chosen to analyze these three accounts in particular because they each rely on a different area of moral psychological data. In addition, each of their theories provides a different way of supporting and challenging *psychological non-cognitivism*. Exploring each of these three accounts will also allow us work through some general problems concerning inferences from the data to conclusions about the plausibility of *psychological non-cognitivism*.

In the coming chapters, I will look to analyse Haidt, Greene and Moll's positions independently. In each chapter, I will look to exposit the author's general theory, briefly present the evidence they offer in its favour, and then consider some objections that can be presented to the theory. The objections I will focus on will both challenge certain features of the author's account, and raise general methodological concerns that impact broader features of this analysis. Addressing these concerns will therefore offer both the benefit of defending the plausibility of the author's conclusions, and of addressing broader concerns that affect the central premises of this analysis. It is important to note that my goal in presenting these positions is not to argue that any of them offer a complete and unproblematic explanation of the psychological processes behind moral judgment. Instead, I will explicitly focus on the common features of these positions in order to reach general conclusions about the psychological nature of moral judgments. With these general points in mind, I will now turn to Jonathan Haidt's Social Intuitionist Model and its explanation of moral judgment.

Chapter 4- Jonathan Haidt- The Social Intuitionist Model

4.1 Exposition

Haidt provides a complex and insightful theory for explaining the nature of moral judgments. Haidt argues that our moral judgments is not a process involving conscious rational deliberation, but a process dominated by a set of fast acting intuitions that operate outside our conscious awareness and control. Haidt names his view the Social Intuitionist Model (SIM)³³ and it looks to explain the relationship between our intuitions and our everyday moral judgments using the connections established by Figure 1.³⁴ This image represents the six links that Haidt's SIM posit are necessary to explain the development and functioning of our moral judgments. I will briefly elaborate on each of these links before moving on to consider the evidence Haidt forwards in favour of his account.

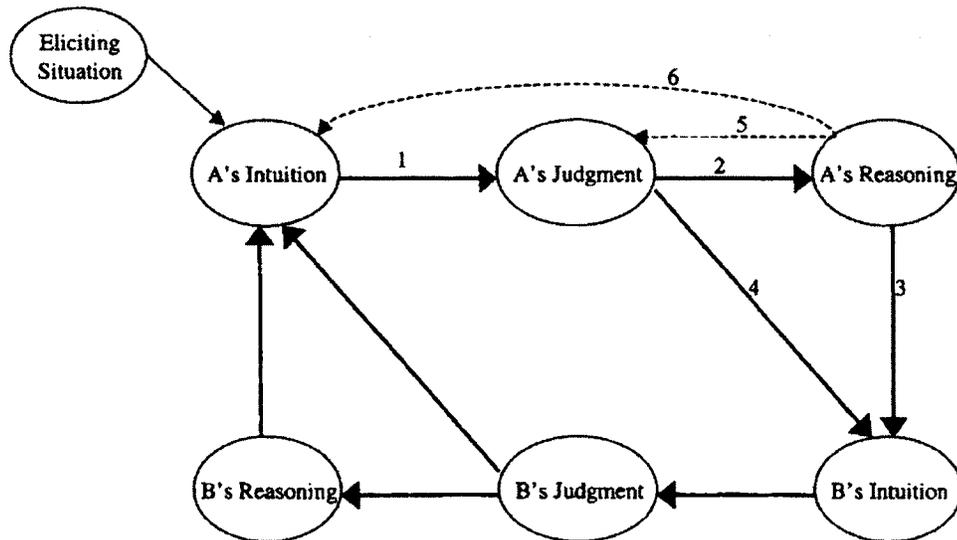


Figure 2. The social intuitionist model of moral judgment. The numbered links, drawn for Person A only, are (1) the intuitive judgment link, (2) the post hoc reasoning link, (3) the reasoned persuasion link, and (4) the social persuasion link. Two additional links are hypothesized to occur less frequently: (5) the reasoned judgment link and (6) the private reflection link.

Link 1 represents the *intuitive judgment link*. The intuitive judgment link in many ways represents the heart of Haidt's theory. Haidt's central observation is that our moral judgments are typically determined by automatic non-conscious entities that he calls the 'intuitions'. Haidt's wording choice is slightly problematic as 'intuition' carries a number of interpretations in the philosophical context. That said, Haidt specifies that by 'intuition' he intends to refer to a kind of mental process which immediately presents us with a particular conclusion without making us aware of the process that lead to this conclusion.³⁵ By defining intuitions in this manner Haidt intends to differentiate two kinds of mental processes: *reasoning* and *intuition*. These processes largely resemble (though importantly aren't identical to) the distinctions we addressed in the last section between cognitive and non-cognitive mental states. The key distinction between these two processes is that reasoning happens consciously and deliberately and takes time, while intuition happens automatically and without the conscious involvement of the agent. The intuitive judgment link therefore claims that our moral judgments appear in our consciousness as a result of these automatic intuitive processes, and not by rational conscious deliberation. This proposal is controversial, as many people believe that their moral judgments are based on some form of rational deliberation, and would be surprised to find that their moral positions are simply the productive of an automatic psychological system operating largely out of their awareness and control. Indeed when most individuals enter into moral debate with one another, they do not generally take themselves to be simply expressing their own intuitions, but to be reasons and arguments for why their position is justified. The next link in the SIM is meant to address this concern and explain how reasoning and rationality is related to moral discussion.

The Second link in the SIM is the *post-hoc reasoning link*. Here Haidt looks to invert our normal understanding of the relationship between our moral reasoning and our affirmation of a particular judgment. According to this link, our moral reasoning isn't responsible for our moral judgments, but these judgments direct and determine this reasoning. Thus, Haidt argues that the reasons we express in favour of our judgments are typically used as a means of defending the conclusions predetermined by our intuitions. Haidt argues that moral reasoning is more often in the role of a lawyer arguing for a particular pre-determined position, rather than a judge or a scientist attempting to reach the position which is best supported by the evidence. In order to make this interpretation of the role of reasoning plausible, Haidt cites research in other fields of psychology that suggest similar post-hoc reasoning is common in other domains of human affairs. For example he cites studies by Nisbett and Wilson that demonstrate the role of post-hoc reasoning in causal explanations.³⁶ In addition he cites the work of Kuhn,³⁷ Kunda,³⁸ and Perkins, Frady and Bushey³⁹ whose findings demonstrate individuals' tendency to automatically present reasons for a particular predetermined hypothesis. Haidt uses considerations from these studies in conjunction with his own observations about moral behaviour in order to conclude that moral reasoning is in the majority of cases, post-hoc.

If Haidt is right in concluding that moral claims are based primarily on our intuition, and reasoning plays no direct causal role in forming our own intuitions, then one might wonder why we bother with moral reasoning at all. If all we are doing is asserting our intuitions, why attempt to ground these intuitions in principles and arguments in the first place? Haidt attempts to answer this puzzle by appealing to the 'social' aspect of the SIM. He argues that the role of rational argumentation in moral

judgment is not to lead the speaker to any new conclusions, but to convince their interlocutor to accept their position. This leads to the next link in the SIM.

This third link in the SIM is labeled the *reasoned persuasion link*. Here Haidt argues that the points being put forward by individual A can have an effect on the intuitions that govern the judgments of individual B. Haidt recognizes that it is relatively rare for an individual to change their mind after being confronted by another individual's moral arguments. Haidt argues that the rarity of these changes in moral opinion is understandable according to the SIM, as an individual's intuition is not based on a particular argument but on the mental processes governing their intuition. As such, identifying a faulty premise or a fallacy in someone's moral beliefs will not thereby convince them that they should change their moral view. Haidt's hypothesis is that most changes in moral opinions occur when the speaker triggers an affectively charged intuition in the listener. By becoming aware of this new affectively charged intuition, the listener might then change their stance on the position. Haidt cites the work of Shavitt⁴⁰ and Edwards and von Hippel⁴¹ whose findings demonstrate the effectiveness of affective persuasion to change affectively based attitudes. Since moral judgments are often affectively based attitudes, Haidt takes this to be reason for supposing that affective persuasion will be the most effective in changing these views as well. While Haidt allows for new intuitions to be implanted in people through moral reasoning and discussion, he does not take it to be the primary method by which individuals develop their moral intuitions. Haidt argues that subtle cues from our social environment do far more to shape our moral intuitions than any argument or rational discussion we might have in our upbringing. This leads to the next link in the SIM.

The fourth link of the SIM is the *social persuasion link*. In this link Haidt posits that the majority of our intuitions are determined by norms adopted by the group we are embedded in. As such Haidt argues that the fact that our friends or allies believe a particular act or behaviour is right or wrong will lead us naturally to have the corresponding intuition. Haidt relies on a number of studies that show that these social factors not only lead to behavioural conformity with group norms,⁴² but a change in privately held judgments as well.^{43 44 45 46} Haidt takes these social forces to be the primary force that determines the content of our moral intuitions. In particular, Haidt believes that different cultures and even different sub-cultures engender the promotion of different kinds of intuitions.⁴⁷

These four links constitute the primary features of the SIM. The solid lines in figure 1 represent the fact that Haidt takes these features to be those that are the most important in explaining our moral judgments. These four primary features are largely automatic and occur with little conscious involvement or control from the agent. The last two processes (5 and 6) represent features of morality that are consciously controlled by the agent. These links focus on the role that conscious rational deliberation can play in our moral judgments. While many individuals believe that conscious rational deliberation plays a significant role in the functioning of our moral judgments, Haidt represents link 5 and 6 in Figure 1 with dotted lines in order to represent the fact that according to his theory, these links play a relatively minor role in standard cases of moral judgment.

The first of these two consciously accessible features of moral judgment is the fifth link in the SIM, *the reasoned-judgment link*. Haidt observes that people may at

times reach a particular moral conclusion through the sheer force of reason and logic alone. One might, for example, observe that one of their higher order moral beliefs (prevent suffering) leads to a conclusion that they had no particular intuition about (the consumption of meat). Haidt argues that while such reasoning does occasionally lead us to new moral judgments, these judgments are generally weaker than those directly backed by an affectively charged intuition. It is weaker in the sense that it is less likely to lead to action and is more easily defeated by competing considerations. In cases where this rational judgment conflicts with an intuition, it leads the agent to have a ‘dual attitude’ wherein the individual’s conscious judgment conflicts with how they intuitively feel about the case.⁴⁸ That said, deliberative reason can at times lead us to a particular moral judgment without an intuition to back it up. While Haidt allows for moral reasoning to lead to a moral judgment without a change in the intuitions, he also allows for cases where our moral reasoning changes our intuition directly. These changes are explained by the next link of the SIM.

The sixth and final link of the SIM is *the private reflection link*. Haidt allows for cases where an agent can change their intuition about a moral judgment through reasoning and rational reflection alone. One of the most common methods for doing so is to intentionally trigger a new set of intuitions through imagining yourself in someone else’s situation.⁴⁹ By imagining yourself in another person’s position you are capable of generating new intuitions based on how the situation is affecting them. Haidt notes that more cognitively focused psychologists such as Piaget⁵⁰ and Kohlberg^{51 52} cite ‘others perspective’ reasoning as one of the primary forms of reasoning which leads us to new moral judgments. While Haidt’s position can accommodate this data, he argues

deliberative reason plays only a minor role in constituting our moral judgments. While individuals can trigger new intuitions by going through the rational process of putting themselves in another person's situation, it is far more common for them to form their judgment on the basis of whatever intuition is triggered by the scenario.

These six links I have considered, connected in the manner presented in Figure 1, constitute Haidt's Social Intuitionist Model. As I mentioned earlier, Haidt's model is a controversial one, as it largely aims to invert our common understanding of the relationship between our intuitions and our moral reasons. Since it is such a controversial proposal, Haidt attempts to make his model plausible by providing a wide array of evidence from behavioural and social psychology that he argues favours his position. This evidence purports to show that automated processes play a far greater role in human cognition than most of us would commonly believe. Such evidence supports Haidt's position insofar as it vindicates the SIM's general strategy of relying on automatic intuitive process to produce our moral judgments.

4.2 Evidence in Favour of the View

Haidt provides four general considerations that he believes help to support his theory of moral judgments. The first set of evidence Haidt presents looks to defend his central premise that "There is a Ubiquitous and Under-Studied Intuitive process at work in much of human cognition".⁵³ Haidt bases his division of the intuitive and the rational spheres of mental processing on a number of well-regarded theories in both cognitive and social psychology. Primarily, he cites the work of Zajonc who advances an affective primacy theory that states that feeling and thought are, to a certain extent, different

systems that function parallel to one another.⁵⁴ Zajonc's theory itself is based on the synthesis of a number of studies from both psychology and sociology. Zajonc infers from the findings of these studies that the affective system is primary as it operates faster than the other system, and is generally more powerful when the two systems conflict. Haidt also cites a number of studies on the 'automatic evaluation effect' which show that very brief and even subliminal presentation of affectively charged words^{55 56}, facial expressions⁵⁷ and photographs of people and animals⁵⁸ “alter the time it takes to evaluate a target object presented immediately afterward indicating that affective processing is at work within a quarter second of stimulus presentation.”⁵⁹ Zajonc's theory and the other observations in social and cognitive psychology support Haidt's assertion that there is a separate and automatic affective process at work in much of our mental activities.

Haidt attempts to add evidence to this position by showing how many other common mental activities appear to function using an automatic non-belief oriented system of evaluation. Haidt observes that one of the emerging views in social cognition is that most of our behaviours and judgments occur automatically “i.e. without intention, effort, or awareness of process.”⁶⁰ First impressions for example seem to be best described as an automatic process rather than a deliberate judgment about an individual. Haidt cites Albright, Kenny, & Malloy who demonstrate that the opinions an individual forms of someone from observing a 'thin slice' of behaviour (as little as 5 seconds) are nearly identical to those they form after longer observation.⁶¹ Individuals also instantly and automatically categorize other people, often attributing them to morally evaluative traits.⁶² These processes show that many of our evaluative processes happen automatically without rational and deliberative involvement.

In addition to our automatic evaluative judgment of others, several studies have shown that our deliberative processes occur in a way consistent with an automatic system. Haidt cites Chaiken's 'heuristic system model of persuasion' which posits that individuals rely on certain heuristics in discussions to conserve cognitive resources, and only employ slower and more systemic processing when there is sufficient complication for the heuristic to not apply.⁶³ One of the heuristics Chaiken identifies is the "I agree with people I like" heuristic. If moral judgments are the automatic processes the SIM takes them to be, then they likely follow a similar heuristic for determining their content. As such, we would expect that when individuals reason with each other about moral issues (links 3 and 4) they are far more likely to change their opinion by reasons presented by their friend rather than the same reasons presented by their adversary. Davis and Rusbully have documented this precise phenomenon, wherein an individual is more likely to agree with an argument if it comes from their allies, a process which they have labeled *attitude alignment*.⁶⁴ In cases where two opponents with mutually exclusive intuitions attempt to argue against one another, it often only leads to further polarization of each other's beliefs. This phenomena is documented by Lord, Ross and Leper and is labeled "attitude polarization".⁶⁵ Haidt argues that these features of our psychological behaviour provide us with grounds for adopting a dual level view of human cognition and this in turn vindicates one of the central components of the SIM.

The second general consideration that Haidt provides in favour of his view comes from a set of evidence that suggests that reason plays a fundamentally post-hoc role in moral judgments. As Haidt puts it, rationality is better categorized as a lawyer defending a client then a judge or a scientist searching for the truth. In order to motivate this

conception of rationality, Haidt provides two examples of types of motivation which heavily bias and direct subject's reasoning. The first are *relatedness motives* and the second are *coherence motives*. *Relatedness motives* refer to those motives that arise from the relationship we have with the individual we are reasoning with. As Haidt puts it, from an evolutionary perspective "it would be strange if our reasoning system was designed principally for accuracy, with no concern for the disastrous consequences of periodically siding with our enemies over our friends."⁶⁶ This consideration helps to motivate the idea that in particular situations reason would be amenable to the more automatic psychological systems that governed social interactions.

One such set of situations that followed this model were observed in a study by Chen, Shechter, and Chaiken.⁶⁷ In this study, the authors demonstrated that when individuals think they are going to be paired up with some one of a particular view, they automatically change their position to fit closer with that individual's view. This sort of behaviour is commonly observed in studies investigating the relationship between social pressure and belief formation. Haidt also observed another set of motivations that subtly affects agents reasoning. Haidt points to data that suggests that individuals often focus on evidence and arguments that support their current position. Haidt refers to these tendencies as *coherence motives*. A long standing history of research on cognitive dissonance demonstrates that individuals often automatically undergo radical changes to their beliefs in order to avoid internal contradictions.^{68 69} More recent considerations from Chaiken, Giner-Sorrolla, and Chen show that this effect is particularly strong in 'self-definitional attitudes and beliefs' such as individuals moral attitudes.⁷⁰

This sort of automatic biasing effect can be seen in studies that challenge individual's moral and political ideologies. Lord et al. found that when subjects were researching the death penalty, they would uncritically accept evidence that supported their position, while critically examining those pieces that went against it.⁷¹ Other studies show how individuals automatically and unconsciously change their beliefs to fit other commitments. For example individuals generally want to believe the world is just and fair and as a consequence when individuals view people who suffer for no reason, they often denigrate or blame them.^{72 73} In addition, Tetlock, Kristel, Elson, Green, and Lerner show that individuals are less willing to use information and counter-factual thinking in particular situations if it could potentially damage their "sacred beliefs".⁷⁴ Haidt argues that each of these observations show how reasoning can be co-opted by automatic psychological systems that look to defend and advance certain conclusions.

This sort of behaviour is also observed in moral judgments. For example, Haidt cites evidence suggesting that when individuals are primed with thoughts of death, they attach themselves more strongly with their cultural worldview.⁷⁵ As a consequence these primed individuals also condemn violations of cultural values more harshly, and provide more praise for moral behaviour.⁷⁶ These observations lead Haidt to the conclusion that human reason is not a mechanism designed to be objective in all cases, but is for the most part designed to fulfill certain social functions by automatically providing justification for our behavior. Haidt doesn't think the upshot of this evidence is that people are stupid or irrational but that we should "look for the roots of human intelligence, rationality and virtue in what the mind does best: perception, intuition, and other mental operations that are quick, effortless, and generally quite accurate."⁷⁷ Haidt thereby concludes that the

more intuitive and automatic elements of our psychology play a primary role in shaping our judgments.

The third general consideration that Haidt provides in favour of his position is the fact that post-hoc rationality is a common feature of human psychology. In other words, Haidt argues that as a rule, we are generally bad at citing the factors that actually influence our judgments. In addition there is good reason to believe that we naturally provide alternative explanations that we then take to be the genuine reasons for our judgment. For example, Nisbett and Schachter conducted an experiment where individuals were asked to subject themselves to increasing levels of electric shock.⁷⁸ Half of the participants received placebo pills while taking the shocks, while the others did not. Those who took the placebo pills misattributed the symptoms of the shock to these pills, and willingly subjected themselves to four times the shock than those individuals who did not receive the pills. When individuals were asked whether or not they were attributing the symptoms to the pill, only 25% of participants were capable of recognizing that this is what they were doing. The remaining subjects provided some alternative explanations to try and explain their tolerance to the shocks.⁷⁹

This data coheres with Nisbett and Wilson's general theory that states that when individuals attempt to explain their actions, they are not trying to accurately represent the cognitive activity that led to their behaviour, as much of this activity is not available to their conscious mind. Instead, they consider the "pool of culturally supplied explanations of behaviour,"⁸⁰ and select the best fitting explanation from this pool. Nisbett and Wilson argue that when we look for an explanation for our action we simply look at whatever pieces of our behaviour fit into the acceptable cultural model and then present those

reasons as our own.^{81 82} Post-hoc reasoning can also be observed in cases where experimenters use hypnosis⁸³ and subliminal presentation⁸⁴ to make people behave in a particular manner. In these cases, individuals actively invent reasons to explain their behaviour. Individuals in these experiments sincerely believe that the post-hoc reasons they are presenting for their actions are actually the reasons for their behaviour. The key observation here for Haidt is that these individuals aren't intentionally lying to cover up their inexplicable behaviour, they are automatically generating plausible stories for their behaviour which they take to be true.

The most compelling evidence for post-hoc rationalization according to Haidt comes from studies of split-brain patients. Split-brain patients are patients which have had their corpus callosum severed in order to reduce the threat of seizures. The corpus callosum functions to communicate information between the left-brain and the right-brain, and so in split-brain patients, these two brain regions operate largely independently of one another. When a split-brain patient moves her left hand by operating the motor cortex in her right brain, the linguistic center in her left-brain automatically provides a story for this behaviour.⁸⁵ This is surprising behaviour, as the right brain motor cortex has no means of communicating with the linguistic center in the left brain. This region of the brain then is seen to automatically provide an explanation for the agent's action, despite its inability to connect with the cause of that action. The linguistic center of the brain so consistently and automatically provides explanations of our action that Gazzaniga has posited it as an "interpreter module" which constantly and automatically provides us with plausible stories for our behaviour.⁸⁶

Haidt argues that the commonality of post-hoc rationality in many features of our lives provides reason to suspect it will be working in moral reasoning as well. In order to bolster this hypothesis, Haidt draws several parallels between the cases that have been discussed and cases of moral reasoning. According to the SIM, when an agent judges something to be wrong, they do so because of an automatic intuitive response that occurs without their control or awareness. While the agent has no conscious access to the reasons for their judgment, they can access the culturally supplied norms for evaluating behaviour and then provide their reasoning using these norms. Haidt labels these norms *a priori moral theories*.⁶ As such, when individuals attempt to provide an explanation for their moral intuition they can appeal to a principle like ‘unprovoked harm is bad’. This norm may not govern the intuitive mechanism that governs their judgment, but it is still a reason they can voice in its favour. Haidt suggests that our *a priori moral theories* will be largely built around these intuitions and so it will appear as if our moral judgments are built from these principles, when in fact it is the other way around.

Reversing the causal relationship between these two features allows us to provide interesting interpretation of a number of observations that are generally meant to favour a rationalist reading of moral judgments. Kohlberg is largely responsible for the cognitive revolution in psychology, a movement which looked to place reason and rationality at the center of explanations of human action.⁸⁷ Kohlberg attempted to explain moral behaviour by dividing groups of individuals into stages of moral development on the basis of the sort of explanations they would provide for why something was wrong. While the theory

⁶ Note that Haidt’s use of phrase ‘*a priori moral theories*’ is entirely distinct from how the phrase is used in general moral philosophy. Here Haidt is only referring to those sets of moral reasons we pick up when we are raised in a particular society.

posited that progression through these stages indicated a sign of moral maturity, Haidt's theory provides an alternative explanation of this progression. According to this explanation, as individuals developed intellectually, they would gain access to new reasons that they could use to justify their intuitive judgments. For example, children in Kohlberg's first stage of development generally explain that an act was wrong because they would be punished for performing it. While Kohlberg generally interpreted this as evidence of an undeveloped moral rationality, Haidt argues that children may provide these explanations because they are reasons that sound plausible to the child, as they are likely reasons that have been presented to them in the past. Post-hoc reasoning then can be shown to explain these more rationally oriented theories of moral judgment. We appear to be reasoning in cases of moral judgments because, as in other social scenarios, we naturally rationalize our action using the cultural information available to us.⁸⁸

The final general consideration Haidt puts forward in favour of his account is the observation that "action covaries with moral emotion more than with moral reason."⁸⁹ In order to help add plausibility to this claim, Haidt cites a number of studies that show that there is a weak link between reasoning and moral action. Before we can consider these studies further, it will be necessary to address some of the considerations that have led individuals to believe that reasoning does play an important role in moral judgment. In particular it will be necessary to explain Blasi's findings that "moral reasoning and moral action are statistically related".⁹⁰ While Blasi only ever viewed the relationship between moral reasoning and moral action as correlational, individuals in the cognitive development tradition argued that this link should be seen as causal. The cognitive development theorists then want to try and preserve the arguably folk-psychological

notion that being better at moral reasoning will lead us to act more morally on account of the fact that this reasoning will lead us to the right moral conclusion more often, and this in turn will lead to action.⁹¹

Upon further investigation, Blasi found that another explanation of the relationship between moral reasoning and moral action was available. According to this explanation, it is not the reasoning itself but the IQ of the agent carrying out the reasoning that predicted moral behaviour. This alternative explanation arguably still favours the cognitive developmental interpretation, as it seems natural to presume that IQ would affect moral reasoning and not some other non-rational psychological system. In response to this, Haidt attempts to provide an alternate explanation for IQ's effect on action. To provide this explanation Haidt relies on Metcalfe and Mischel's dual process theory of willpower.⁹² According to this position, there are two systems that govern behaviour in cases of temptation. The 'hot' system runs on quick emotional responses and relies on amygdala-based memory. The 'cool' system is designed to deal with “complex spatiotemporal and episodic representation and thought.”⁹³ It employs hippocampal memory and relies on the activation of the frontal lobe.

This theory was developed in part to explain a number of findings. In particular it was motivated by the surprising discovery that the number of seconds a preschooler delayed an immediate reward (one marshmallow sitting in front of them) for a larger future reward (two marshmallows) was a strong predictor for their cognitive competence 13 years later (as measured by their SAT scores and the ability to demonstrate self-control).⁹⁴ These findings support the hypothesis that certain children start off with stronger 'cool' systems as a result of better or faster frontal cortex development.⁹⁵ A

stronger frontal cortex leads these individuals to exhibit more intelligence and willpower later in life, resulting in better SAT scores and better performance on tests measuring willpower. The fact that intelligence and willpower are both a result of a strong frontal cortex can in turn explain why moral reasoning and moral behaviour are correlated. According to these observations, individuals with higher moral reasoning will also have a stronger inhibitory system, and as a consequence will be better at inhibiting immoral behaviour.

This prediction is consistent with the longstanding observation that there is an inverse relationship between criminality and IQ (even after isolating socio-economic factors).⁹⁶ It is important to note that what has been said so far only applies to limiting immoral behaviour, not to promoting moral behaviour. When it comes to positive moral motivation, the link between intelligence and moral behaviour becomes less clear. For example, several studies examined pro-social moral exemplars with non-exemplars and showed that the two groups did not differ in their moral reasoning abilities (as measured by Kohlberg and Rest's scales).^{97 98} A more comprehensive review of the standard tests conducted for moral reasoning showed that while there was a relationship between these scores and negative moral behaviour, there was no relationship between the scores and positive moral activity.⁹⁹

These considerations lead Haidt to the conclusion that moral reasoning does not in itself provide us with motivation for moral action. After establishing this point, Haidt attempts to show that emotions play the primary role in providing this motivation. Haidt derives most of his evidence of the role of emotions in moral motivation from the study of psychopaths. Haidt takes psychopaths to be ideal subjects for this analysis because

they are individuals who are capable of advanced higher-order reasoning, but appear to lack certain important moral emotions. Cleckley's study of psychopathy shows that psychopaths suffer from an inability to generate a number of affect reactions, "particularly those that would be triggered by the suffering of others (remorse, sympathy), condemnation by others (shame, embarrassment), or attachment to others (love, grief)."¹⁰⁰ The emotional deficits in psychopaths are then thought to be responsible for their antisocial behaviour.

Several studies also suggest that this behaviour is due to selective damage in the prefrontal cortex.¹⁰¹ In particular Haidt relies on the research done by Damasio and colleagues who found a consistent set of anti-social personality changes that accompany damage to the ventromedial prefrontal cortex (VMPFC).¹⁰² Individuals with this damage are fully capable of retaining information about moral rules and social conventions, and are even capable of performing well in hypothetical moral dilemmas. Despite the fact that these individuals maintain the rational capacity to think about morality in theory, they lose the ability to be moral in practice. Damasio goes on to show that the central capacity that is lost from damage to the VMPFC is the ability to have normal emotional responses to the world. When measuring the skin conductance of subjects exposed to emotionally provocative photos (nudity, mutilation, people dying), normal subjects show increased receptivity, while those with damage to their VMPFC experienced no response.¹⁰³ This same behaviour was observed in individuals with psychopathy.¹⁰⁴ The subjects with VMPFC damage in these cases report that they know that they should feel something in response to the pictures but they simply feel nothing. For Haidt's purposes, these cases are important because they show that moral reasoning can stay intact, while

moral motivation is lost. This shows that moral reasoning cannot be a sufficient condition of moral judgments, as moral judgments are generally also thought of as having a motivational/practical component to them.

After showing that moral reasoning isn't enough to provide us with moral motivation, Haidt goes on to review the evidence that suggests that the emotions play an important role in providing us with this motivation. Haidt briefly reviews the debate surrounding the empathy-altruism hypothesis which states that empathy aroused by the perception of someone else's suffering produces altruistic motivation to reduce that suffering.^{105 106} While there is a live debate over the plausibility of this view, both its proponents and its opponents agree that the motivation to help others comes from a primarily affective system which includes empathy, as well as sadness, guilt and shame.¹⁰⁷ Thus Haidt concludes that the primary motivation for moral action is from the automatic non-rational domain of the emotions and not from the rational domain of reflective conscious deliberation.

4.3 Criticism: Saltzien and Kasachoff

Before considering some of the major criticisms of Haidt's SIM, it will be helpful summarize the evidence he has provided in its favour. Each of the pieces of evidence Haidt presents is meant to add to the plausibility of certain features of his position. In particular Haidt wants to make two general claims: 1) that there is an affective intuitive system at work as the dominant system in much of human cognition and 2) that this system, and not the slower more rational one, is primarily responsible for moral judgments. Haidt is also committed to the more specific claim that these judgments

operate through the six links posited by the SIM. When objecting to Haidt's position one can take issue with any of these three claims. In objecting to the first claim, one could argue that our psychological systems are both more unified and more rational than Haidt's model would allow. In objecting to the second claim, we could remain ambivalent about the role of rational conscious judgment in more general cognition, but insist that in the moral domain, conscious rational judgment is more important than the SIM would have us believe. Conversely one could agree with much of the evidence Haidt advances in favour of his position, but argue that a different model could better accommodate this data. Along a similar line, one might pursue a line of general scepticism toward Haidt's position and argue that the data he presents fails to vindicate his particular theory, and that the same data could be interpreted to different ends.

The objection I will focus on the second general line of argument, namely that reason plays a more important role in moral judgment than Haidt allows. The critique I will consider speaks to Haidt's position directly, offering particular reasons for why his affect-centered account is inadequate. For the purposes of this project, it will be most beneficial to focus on this line of criticism as it more directly speaks to issues which affect the plausibility of *psychological non-cognitivism*. While many features of Haidt's theory may be open to reasonable critique, for the purposes of my project these more particular criticisms are not as relevant as the more general point that automatic and seemingly non-propositional mental states play a large role in our moral judgments. As such, I will focus on Saltzstein and Kasachoff's critique that these automatic social-intuitive features are unable to provide an adequate explanation of moral judgments.

The first major critique to Haidt's position that I will consider comes from Herbert Saltzstein and Tziporah Kasachoff 2004 article *Haidt's Moral Intuitionist Theory: A Psychological and Philosophical Critique*.¹⁰⁸ In this article Saltzstein and Kasachoff (S&K) make three general criticisms against Haidt's position. First, they claim that Haidt's position reduces social influence to compliance while ignoring other more rational forms of influence. Second, they argue that this position fails to distinguish between the causes that lead to the development of morality, and those that operate in its deployment. Thirdly, they argue that Haidt's position fails to distinguish between causes and 'reasons' or justifications.

S&K's first objection rests on the observation that there are many forms of influence that are more rational than those that Haidt's model accepts. S&K argue that Haidt's account does not allow for our intuitions to change as a result of reasoned argument and persuasion. Intuitions for Haidt's position are the starting point for moral judgments, but S&K claim that in moral discourse our moral intuitions are often generated or changed as a result of moral discussion. Indeed in certain debates we may have no moral intuition to begin with, and then generate new intuitions as a result of reasoned deliberation. In addition, there are cases where we have a strong moral intuition against killing an innocent individual, but we can override that intuition if it will, for example, result in saving the lives of a number of people.¹⁰⁹ There are other cases as well, where subjects might find themselves with an immediate repulsion to something (such as homosexual activity) but still not judge these actions to be morally wrong. In addition we also want our moral judgments to be consistent with one another, and this

desire for consistency is more consistent if we interpret our judgments as beliefs rather than mere intuitive preferences.

As a more general point, S&K observe that Haidt fails to sufficiently explain how moral discussions influence our moral intuitions. They recognize that Haidt is committed to the claim that rational discussions can only cause individuals to change their mind by triggering new non-rational intuitions. S&K challenge this assertion by citing a number of theories which suggest that there are different ways which our interactions with others can change our position.¹¹⁰ Some forms of persuasion are social and non-rational while others are deliberative and reflective. S&K argue that Haidt fails to appreciate the diversity and pervasiveness of this more reflective and deliberative form of social influence, and argue that individuals change their mind on the basis of reasons far more than Haidt's model will allow.

The second objection S&K present against Haidt's position is that it fails to distinguish the development of morality from its deployment. In particular S&K want to argue that Haidt's observation that moral judgments rely on automatic responses does not rule out the possibility that they were at one point based on rational deliberation. S&K observe that many processes we establish through conscious reasoned steps become habituated over time. Mundane examples of this include processes like learning to tie our shoes or learning the rules of addition. Along these lines, our moral attitudes might have been established through reasoned deliberation, but now have become automatic due to habituation. As such in order to make his position plausible Haidt has to do more than merely show that our moral responses are automatic, he must also eliminate the possibility that these automatic responses are simply habituated rationality.

The third objection presented against Haidt's account is that it fails to distinguish between causes and reasons for action. Recall that in defending his position, Haidt argues that we do not have conscious access to the psychological processes that cause our moral judgments, and as such rationality only plays a post-hoc role in justifying the outputs of these non-conscious systems. S&K argues that this line of reasoning commits Haidt to a conflation between the justificatory reasons we have for an action, and the psychological explanations for why we acted the way we did. S&K put the point as follows:

“If we are trying to show that a moral judgment is correct, appeal must be made to criteria that morally justify the judgment and not to processes that made it occur. If you want to know whether what I did was morally correct, you are asking not what made me do what I did but whether what I did was the right thing to do in the circumstances”¹¹¹

The idea then is that moral judgments are concerned not with the mental state that caused us to have that judgment, but the reasons we have for making that judgment.⁷ If this is correct, then Haidt's analysis has made a fundamental error. If moral judgments are really about the reasons we can express for those judgments, then the SIM could at best show us that a series of complicated automatic systems skew our perspective and distract us from the proper focus of our moral judgments. This is an important point to consider, not just for Haidt's position but also for our more general discussion concerning the status of the plausibility of *psychological non-cognitivism*. If moral judgments are only

⁷ It is important to note that our reasons for action can also be the cause of this action. When we act for a reason, that reason is also the psychological cause of our action. Thus while we generally discuss reasons independently of their role in causation, it is important to be aware of the fact that reasons can also be causes.

concerned with the reasons we can voice in their favour, reasons that we process and generate in a conscious and deliberative fashion, then it seems that *psychological non-cognitivism* would be necessarily false, as anything that was a moral judgment would be a reason, and reasons are represented by cognitive mental states. Thus it will be important both for Haidt's position, and for the broader project, that we consider Haidt's response to these objections.

4.4 Response to Saltzstein & Kasachoff

Haidt provides a detailed response to S&K in his article "The Emotional Dog Gets Mistaken for a Possum."¹¹² Haidt chose this title to draw attention to the fact that he takes many of S&K's comments to be based on an incomplete reading of the SIM. As such, much of Haidt's response revolves around pointing out those features of the SIM that S&K failed to consider in their critique and then showing how these unconsidered features counter their objections. This is particularly the case in regards to S&K's first criticism. Haidt argues that this criticism is based on a failure to appreciate the role of the *reasoned persuasion link* in his model (link 3 in Figure 1). Recall that according to this link, one of the major ways individuals' intuitions change is by having other individuals in their community reason with them about their judgments. Haidt's position therefore allows for moral discussion to lead to changes in our moral judgments, he simply argues that these changes are not necessarily attributable to any sort of conscious reasoning by the agent involved. While he doesn't rule out this happening on occasion, he argues that the change of opinion is far more often attributable to individuals being stimulated to have an intuition that they did not possess before. Thus Haidt can accommodate S&K's

assertion that there are two distinct ways to influence individuals moral positions, one social and one more rational, by appealing to the distinction between the *reasoned persuasion link* and the *social persuasion link*.

S&K cite our desire for coherence amongst our judgments as a reason for favouring a more rational construal of our moral judgments, but this desire for coherence can be accommodated in Haidt's account by positing it as an innate feature of our intuitions. Given that the social intuitions are meant to coordinate social activity, it makes sense that they would be structured in a way that would favour coherence, as an inconsistent set of moral intuitions would lead to social disarray. Thus we see that much of S&K's first objection to Haidt can be attributed to an incomplete and uncharitable reading of the SIM. While many features of their first objection rest on a misreading of SIM, Haidt also looks to address a substantive disagreement that ties many of these criticisms together. This substantive point concerns the extent to which conscious deliberation leads to a change in moral option.

S&K maintain that moral reasoning works more or less as it appears to, we argue about particular claims, and then when one is confronted with a stronger argument one changes their mind. The main focus of Haidt's project is to provide an alternative story to this commonly held view. As we have seen, Haidt relies on evidence forwarded by Edwards and von Hippel¹¹³ and Shavitt¹¹⁴ that shows the importance of affect priming in changing individual's affectively based attitudes. This investigation, as well as many others like it shows us that human reasoning is not always what it seems, and that our psychologies preform many quick and subtle changes in response to affect laden stimuli. Affective priming has been shown to have a substantial effect on changing individuals

affectively laden attitudes, and since in the majority of moral judgments are affect laden, Haidt concludes that those arguments which appeal to an affective component will have a far greater effect on shaping an individual's moral intuitions. Given these considerations and the others discussed in the previous section, Haidt concludes that the non-rational features of moral arguments will have a much greater effect on our moral beliefs.

It is important to note that Haidt does not require all changes in moral opinion to be based on these affectively charged arguments. The *reasoned judgment link* allows for the possibility that an individual employing conscious effort and logical reasoning can alter their judgment. This in turn allows for the possibility of the individual changing their beliefs on the basis of reason alone. While this is possible under Haidt's model, his substantive point is simply that this is not what takes place in standard cases of moral judgment. Haidt's position is therefore not threatened by S&K's first criticism, as it can accommodate the general concern while providing an alternative account of how moral reasoning affects our future judgments.

The next general criticism S&K advanced was the claim that Haidt's position fails to consider the possibility that moral judgments are automatic in their application, but were originally established through reasoned argument. I argue that Haidt can respond to this criticism simply by appealing to his own explanation for how our moral intuitions get established. I did not discuss this feature of Haidt's theory earlier, but roughly, Haidt believes that our intuitions are established both biologically and through particular features of cultural conditioning. In his more recent renditions of the SIM, Haidt narrows down 5 general categories of moral intuitions: harm, fairness, ingroup-outgroup,

authority and purity.¹¹⁵ In this work Haidt argues that these categories are set up as a 'first draft' in our biology and then get refined and focused through cultural conditioning. Each cultural has a set of practices that helps to set the parameters of those situations that will trigger our moral intuitions. We observe these practices starting at a very young age, and through our observation of the norms of others, we internalize a particular set of moral intuitions. If this story is plausible, then Haidt rules out the notion that our moral judgments could be established by habituated rational deliberation.

While a full discussion of the plausibility of Haidt's explanation of moral development would take us far from the task at hand I argue that even without this investigation, there are a number of psychological considerations that favour Haidt's proposal over S&K's. First of all, many moral intuitions are expressed by children long before they have developed enough to do any sort of complex moral reasoning.¹¹⁶ This seems to call into question S&K's proposal, as these children clearly were not capable of cognizing the sort of conscious rational deliberation necessary to form a critical moral judgment. That said, S&K might claim that these judgments are rationally assented to later in life, and that this rational assent leads to the automaticity of moral responses through a kind of habituation. This sort of response though seems largely post hoc, as it is clear by studies of moral responses in children that the psychological activity necessary for the expression of moral intuitions develops faster, and independent of, the activities of higher order rational thought. Haidt's position can respond to S&K's criticism therefore by providing a plausible explanation for the automatic nature of moral judgments.

The third and most interesting objection provided by S&K concerns Haidt's alleged conflation between explanations and justifications for moral judgments. S&K

argue that the focus of our moral judgments are the reasons we can give in their favour, not the psychological mechanism that caused them to occur. Haidt's response to this objection begins by revisiting what he takes reasoning to be:

“After reviewing the literature on everyday reasoning, I concluded that the critical feature of reasoning is that it occurs in sequential steps, at least a few of which must be performed consciously. I defined moral reasoning as “conscious mental activity that consists of transforming given information about people in order to reach a moral judgment” ”¹¹⁷

Thus for Haidt, in order for moral judgments to be primarily concerned with what is rational, they would have to be constituted by several sequential steps which occur consciously. Haidt observes that S&K define rationality in a far broader manner. They seem to define rationality “as any mental process that is responsive to reasons and reason giving.”¹¹⁸ Haidt argues that this conflates two very different cases. The first case has two individuals discussing a moral issue, and in this discussion person A understands B's contrary position, decides it is correct and then changes her mind. For Haidt, as long as there was two consciously understood sequential steps involved in A changing her mind, Haidt fully accepts that the change in A's judgment is a product of conscious reasoning. The other case could involve the same conversation, but in this case A doesn't go through conscious reasoning but surely just “gets” it. This for Haidt is not a case of reasoning, but a case of an intuition. In these cases, Person A has not gone through any moral reasoning herself, but has had a new intuition triggered by the words of Person B. Haidt then allows for reasons and reasoning to influence behaviour in a way that is, by his

definition, not rational. Reasoning for Haidt is a conscious system of deliberation, and therefore anything relegated to unconscious automatic cognition is not 'reasoning' in a substantive sense. Haidt's first point against S&K's objection then is that it is a mistake to label all systems that are reason responsive as 'rational', as some systems change in response to reasons without being rational.

The other more general, and for the purposes of this project, more interesting feature of S&K's criticism is the idea that the proper focus of moral judgments are reasons. As they put it, if one were to respond to the question 'why do you think that is wrong' with 'because my psychological system x did y' we would accuse them of failing to understand our question. This objection is important for my more general project, because if S&K can show that the proper concern of moral judgments are reasons, and these reasons are expressed propositionally, then it seems to lead to a necessary refutation of *psychological non-cognitivism*. That said, I do not think that S&K can plausibly assert that the psychological states relevant to moral judgments can be limited to the reasons we provide for and against them. Indeed, one of the biggest insights of Haidt's system is that our moral reasoning is only one part of what determines the content of our moral judgments. When an individual makes a moral judgment, they are not merely adjudicating amongst reasons; they are reacting to affectively charged intuitions that play a large role in constituting their judgment. S&K might press the point and claim that these features should only be seen as influencing our moral judgments, and not in any way constituting them. I believe that this move would be unwarranted, as according to Haidt's system, our intuitions don't simply motivate our judgments, they also constitute an important feature of these judgments.

Our moral reasoning is built on certain intuitions about what is right and wrong. Thus in order to make sense of moral judgments as they operate in normal subjects, we ought to include this system that does so much to provide the motivational and affective content of our moral judgments in the first place. It is of course possible for us to ignore the role these intuitions play in our moral judgments and focus exclusively on the moral reasoning, but doing so would fail to accurately represent the psychological states involved in our moral judgment. Since *psychological non-cognitivism* is a thesis that is concerned with the nature of these mental states, we cannot rightly exclude the intuitions from our analysis. With this consideration in mind, Haidt is able to resist S&K's third line of criticism. For those readers who are interested, I have considered an additional set of objections to Haidt's account by Pizzaro and Bloom in Appendix A. While these objections do bring out some interesting features of Haidt's position, it is not essential that we consider them in order to advance the primary focus of this work. Thus in the interest of continuing our central investigation, I will now set out to consider the effects of Haidt's position on the plausibility of *psychological non-cognitivism*.

4.5 SIM's Implications for Psychological Non-Cognitivism

Now that we have analyzed in detail the main features of Haidt's position, we can go on to consider the implications of his view on the plausibility of *psychological non-cognitivism*. While a full analysis of the implication of Haidt's position will have to wait until after I considered Greene and Moll's accounts, on the basis of what has been seen so far I believe I can safely claim that Haidt's SIM favours *psychological non-cognitivism*. There are a number of features of the SIM that support *psychological non-cognitivism*.

First and foremost, Haidt's conception of an 'intuition' is very close to the definition of a non-cognitive mental state which we discussed in the last sub-section. As we have seen, Haidt's intuitions function both automatically and without the agent's direct conscious involvement. As such, intuitions clearly satisfy the two secondary features of non-cognitive mental states identified in the previous chapter. The question then becomes whether or not the 'intuitions' as Haidt describes them, have propositional content, and whether they can be properly categorized as 'beliefs'.

While Haidt never explicitly evaluates the intuitions within the context of propositional content, I think it is safe to infer that the intuitions are largely non-propositional entities. First off, Haidt does quite a lot to tie the intuitions to affectively charged psychological systems. As we discussed earlier, affectively charged states are prime candidates for non-propositional mental states, as significant features of these mental states are neither truth-apt nor concerned with the predication of properties. The fact that intuitions rarely operate in moral reasoning, but are instead only indirectly affected by the reasoning provided by others indicates that they are not sensitive to the propositional content that dominates conscious rational deliberation. The intuitions' insensitivity to these propositional beliefs gives us further reason for thinking that they are non-propositional. I argue that these features of Haidt's intuitions ought to lead us to the conclusion that they are non-propositional mental states.

Given the intuitions' prominent role in Haidt's theory of moral judgment, this might lead us to the conclusion that Haidt's theory strongly supports *psychological non-cognitivism*. That said, it is important to keep in mind that the SIM incorporates features other than the intuitions in its explanation of moral judgments. Haidt's system does allow

for moral reasoning to both directly affect the output of our moral judgments, as well as change our moral intuitions in particular situations. The puzzle then is to determine to what extent these other features ought to change my previous assertion that Haidt's account favours *psychological non-cognitivism*.

In order to properly address this problem, we need to revisit the distinction made in the last chapter between those features that constitute moral judgments and those that merely influence them. Given the different roles that reasoning and the intuitions play in Haidt's account, it is difficult to determine precisely which features should be viewed as constituting our moral judgments, and which should be viewed as merely influencing these judgments. Recall that reasoning does play an important role in several 'links' of the SIM including the reason persuasion link and the private reflection link. Haidt also fully recognizes that moral judgments are often presented in the form of propositional claims and arguments. As such, a tempting interpretation of Haidt's position, shadowing one of our considerations from S&K, is that it postulates that moral judgments are essentially propositional, and that these propositional beliefs are merely influenced by our non-propositional intuitions. A proponent of this interpretation of Haidt's position would bolster this position by arguing that if moral judgments were fundamentally non-propositional, then the sort of reasoning outlined in the reasoned persuasion link of Haidt's model would be impossible, as non-propositional content cannot enter into acts of moral reasoning. Since moral reasoning is possible by Haidt's model, our moral judgments must be, in some sense, propositional.

The easiest way to accommodate this observation with the important role played by non-propositional intuitions would be to hold that the intuitions subtly influence our

moral judgments, but that when we form the judgments, their essential mental content is propositional. While this is an interesting interpretation of Haidt's position I believe it is mistaken. I argue that it is mistaken because it fails to appreciate the degree to which the intuitions not only *influence* but also *constitute* the content of our moral judgments. For Haidt, the intuitions help to explain not only our disposition to form a particular moral judgment, but the character and motivational force of the judgment itself. It is precisely the inclusion of the intuitions that helps to delineate a moral judgment from some other normative requirement. So while Haidt's account allows for moral judgments to play a role in rational deliberation, he allows for the intuitions, which I have argued are non-propositional, to play an important constitutive role in what it is to have a moral judgment. This observation though leaves us with a new puzzle as to how the non-propositional features of an intuition could play a role in cases of moral reasoning, insofar as reasoning generally only involves propositional content. In order to address this puzzle, we need to consider how non-propositional mental states might come to play a role in normal cases of propositional reasoning.

In order to consider how a non-propositional mental state like an intuition can play a role in reasoning it will be helpful to look at how other non-cognitivists address this sort of problem. When an individual advances a non-cognitivist position about a particular region of discourse it is important to recognize that they are not arguing that the sentences expressed in this discourse lack any determinate meaning. Instead, they are only arguing that this meaning originates from mental states which are non-propositional. In other words, the meaning of the discourse comes from a feature that is not truth-apt and which is not designed to predicate properties. Each non-cognitivist account provides

a different explanation for what constitutes the meaning of these concepts. Once this meaning has been defined by the appropriate non-propositional feature, the agent can then make meaningful claims about the object in question. While this explains how non-propositional concepts can be meaningful to the agent, it still does not explain how these meaningful claims can be used in seemingly propositional forms of reasoning. While many different non-cognitivists have adopted different strategies to address this problem, for the purpose of this project I am going to focus on a strategy which attempts to solve this problem by having us represent propositionally the meaningful features of our non-propositional mental states.

According to this strategy, a non-propositional area of inquiry is one wherein the central meaningful features of the concept are provided by non-propositional mental states. Once the meaning of these concepts are provided by the non-propositional mental state, the agent can then represent the fact that they are having these mental states in a propositional manner. This process would translate roughly to an agent 'A' having a non-propositional mental state NP, and then representing the fact that they are experiencing NP to themselves as a separate propositional-belief state (A(NP)). Thus A can have the propositional belief A(NP) while still allowing for the non-propositional features of NP to constitute a substantive feature of the concept in question.⁸ By referring to the fact that we are experience NP, we can translate our non-propositional mental states into a propositional form that can then enter into conscious rational deliberation. So returning to Haidt and the intuitions, the meaningful content of our moral judgments could arise from an expression of our intuitions, which we could then represent propositionally in a

⁸ A similar strategy is adopted by Jesse Prinz in *The Emotional Construction of Morals*.

manner that would enable us to use this information in standard forms of reasoning. This sort of general approach could explain how the non-propositional features of moral judgments could come to play an important role in propositional moral reasoning. The question then arises, if this strategy is successful, to what extent does Haidt's position support or refute *psychological non-cognitivism*? While a full exploration of this question will have to wait until after I consider the evidence from Greene and Moll's accounts, it will be sufficient for the time being to note that both rational and intuitive features seem to play an important role in Haidt's theory. That said, intuitions clearly play a more important role, and if the argument I have just presented is plausible, we have reason to believe that these non-propositional mental states are the primary constitute features of our propositional moral judgments as well. This fact seems to lend support to *psychological non-cognitivism*, though questions concerning the extent to which it vindicate this thesis will have to wait until we consider the rest of the data. To this end, I will now move on to present another contemporary moral psychologist whose position largely builds off of the data Haidt used to support his position.

Chapter 5- Greene's Dual-Process Model

5.1 Exposition

Joshua Greene proposes a model of moral judgment which is distinct from the SIM but shares several important features.¹¹⁹ First, like the SIM, Greene's theory of moral judgment posits that two psychological systems operate in moral judgment, one which is slow, conscious and deliberative, and the other which is quick, unconscious and automatic. Unlike Haidt, Greene divides these two systems along different categories of moral judgments. He argues that the quick and automatic system is one largely governed by our emotional responses, and it leads us to form deontic moral judgments. Deontic moral judgments are concerned with questions of rights and duties, and are roughly judgments that evaluate things to be right or wrong in themselves, not as a product of their future effects. Contrarily, our slow rational and deliberative judgments produce consequential moral judgments. These judgments view the moral value of an action on the consequences of that action alone. Green's dual process view then attempts to explain many cases of internal moral conflict as the product of these two different cognitive systems conflicting with one another.

Greene's position originally emerged in part to explain the radically different responses generated by two thought experiments. The first experiment is the trolley problem. In the trolley problem, subjects are asked if they would throw a switch to divert a trolley from a track with five people on it to a track with only one individual, effectively killing one to save five. In this case, most subjects will choose to divert the trolley, saving five and letting one die. This original scenario is labeled the *switch* scenario because the subject is diverting the trolley using a switch located a safe distance

away from the gruesome event. The other case is the *footbridge* dilemma. In this example, there is a trolley heading for five individuals on the track, but this time your only means of stopping it is to push a fat individual standing on a footbridge onto the tracks, thereby stopping the trolley. In this scenario, the majority of individuals chose not to push the fat man off the bridge in order to save the five on the track. The puzzle then is to explain why individuals chose to sacrifice the one to save the five in the *switch* scenario but not in the *footbridge* scenario. Greene's Dual-process theory of moral judgment looks to explain this difference by positing that each triggers a distinct psychological system responsible for the judgment. The *switch* scenario involves the engagement of the slower deliberative psychological system that appeals to consequentialist considerations. As such, this view concludes that one ought to throw the switch because the overall harm will be reduced (one death rather than five). The *footbridge* scenario on the other hand is thought to trigger the automatic deontic moral response in the agent. In this case, the individual is overcome with the automatic negative emotional response that comes with the idea of pushing someone to his or her death. This automatic negative feeling leads the individual to conclude that pushing the fat man off the bridge is morally impermissible.

The question then is why does the rational-consequential system become activated in one scenario and the emotional-deontological system become activated in the other? Green believes that the relevant distinction between the two scenarios is that the *switch* scenario is 'impersonal' while the *footbridge* scenario is 'personal'. Greene defines 'personal' scenarios as those which involve "(a) serious bodily harm, (b) to a particular person or group, where (c) the harm does not result from deflecting an existing threat."¹²⁰

His position therefore holds that those situations that have these characteristics will likely trigger our automatic and emotive moral response, while those that are more impersonal will trigger our cooler more reflective rational judgment. Much like Haidt's theory, Greene's position is controversial, and so it will be important to consider the evidence he puts forward in its favour.

5.2 Evidence

The first major set of evidence Greene provides in forward of his view originates from his original publications proposing the dual-processes theory.¹²¹ In this publication, Greene and colleagues presented subjects with a battery of moral scenarios while they were connected to devices that would collect functional magnetic resonance imaging (fMRI) and reaction time (RT) data from their responses. These scenarios were divided into 'personal' and 'impersonal' scenarios along the lines discussed earlier. Personal dilemmas included the *footbridge* scenario, a case of stealing one individual's organs to save five other individuals lives, and a case of throwing some people out of a lifeboat in order to save the rest. The more impersonal dilemma's included the *switch* scenario, a case considering keeping money from a lost wallet, and a case of voting for a policy which is anticipated to have more deaths than its alternative.

In the first experiment (Experiment 1) individuals were exposed to a set of 60 scenarios. The fMRI data from these questions showed that distinct regions of the brain were activated in 'personal' and 'impersonal' scenarios. This data showed that in personal scenarios, regions of the brain associated with emotion were active. In addition areas with working memory were shown to be less active in these scenarios, a behaviour

consistent with observations of normal cases of activation of emotional regions of the brain. Experiment 2 replicated the materials from Experiment 1, but this time tested for RTs. Greene's theory predicted that there would be longer response times when individuals responded "appropriate" to 'personal' moral dilemmas as such a response would require the agent to override their initial emotional response with a more deliberative one. These predictions were largely vindicated by the data. In impersonal dilemmas, there was no significant difference in RT between responses of 'appropriate' and 'inappropriate'. That said, either choice was slower overall than judgments of 'inappropriate' in personal scenarios, showing that emotional responses happen quicker overall than their 'rational' counterparts. In addition, by correlating results between the fMRI and the RT tests, Greene observed that regions in the dorsolateral prefrontal cortex (DLPFC) were activated in high RT-trial. This, he argued, reflected the engagement of abstract reasoning processes and cognitive control.

The evidence generated by Greene et al.'s initial investigation lead to further experiments designed to evaluate the plausibility of the dual-process view. The first study to note is Greene et al.'s 2004 paper *The Neural Bases of Cognitive Conflict and Control in Moral Judgments*.¹²² This study looked to show that the brain regions associated with abstract reasoning and cognitive control are recruited to resolve dilemmas where utilitarian and more emotional evaluations conflict. To test this aspect of their position Greene and colleagues came up with a set of dilemmas that would bring these two systems into conflict. A good example of one such scenario is the *crying baby* dilemma. In this scenario, enemy soldiers have taken over your village and their orders are to kill all the villagers. Your baby starts crying, and the only way to make it stop is if

for you to smother it to death. If you do not do this, the soldiers will find you and kill you, your baby and the rest of the survivors. It goes without saying that this is a difficult moral dilemma, as either option leaves one with the impression that one has done something wrong.

When confronted with this scenario, subjects tend to respond slowly and exhibit no consensus in their judgment. This dilemma, like the others used in the 2004 study is structured to put aggregate welfare in direct conflict with a 'personal' moral violation. In these cases Greene predicted that the anterior cingulate cortex (ACC), and areas of the brain associated with cognitive conflict, will be activated. Green also predicted that the control-related processes in the DLPFC will activate and tend to favour a utilitarian conclusion for the scenario. Greene contrasted these hard cases with 'easy' personal dilemmas that lent themselves to rapid and uniform judgments amongst subjects. An example of this case would be the *infanticide* dilemma wherein a teenage mom must decide whether or not to kill her unwanted infant. In these cases the strong negative social-emotional response associated with killing the child overwhelms the non-existent cognitive justification for the action.

Greene's first analysis tested the reaction time (RT) differences between cases like the *infanticide* and *crying baby* dilemma. He predicted that RTs in the harder dilemmas would increase due to the mitigating effects of the DLPFC. In his second analysis, Greene had individuals answer whether the utilitarian option in the hard moral dilemmas was 'appropriate' or 'inappropriate'. This experiment looked to isolate those regions of the brain that were responsible for the subject producing a utilitarian response in these cases. The results from these two analyses largely confirmed Greene et al.'s initial

hypothesis. Analysis 1 showed that individuals faced with difficult moral dilemmas had both a higher RT as well as a greater activation of the DLPFC. The results from Analysis 2 demonstrated that individuals who select ‘appropriate’ in the difficult moral dilemmas activated a number of distinct brain regions from those who selected ‘inappropriate’. These regions included certain specific areas of the DLPFC as well as the ACC, both of which are associated with cognitive control.¹²³

Together these two analyses provide evidence for Greene’s dual-process theory, as they show that cognitive conflict arises in cases where utilitarian reasoning and emotional responses clash. In addition these results demonstrated that utilitarian reasoning requires more time and requires activation of the DLPFC, and that in cases where emotional and utilitarian judgments conflict cognitive control processes in the brain need to be activated in order for the utilitarian judgment to be made. This empirical data reflects Greene’s prediction that two distinct systems are operating in moral judgments, and that when they conflict, the agent has to exert some cognitive control in order to reach the rational-utilitarian conclusion.

The last piece of experimental evidence I will consider in favour of Greene’s account comes from his 2008 article *Cognitive Load Selectively Interferes with Utilitarian Judgment*.¹²⁴ In this article Greene and colleagues attempt to add further plausibility to their position by testing the effects of cognitive load on moral judgments. Since the dual-process theory posits two distinct systems of moral appraisal, we should expect each system to be affected by cognitive load in a different way. In particular we would expect the more cognitively demanding system to be effected by an increase in cognitive load, while the automatic system would remain relatively unaffected. Greene et

al. attempt to test this prediction by presenting participants with “high-conflict”¹²⁵ ‘personal’ moral dilemmas. These included the *footbridge* dilemma as well as more difficult dilemmas like the *crying baby* dilemma. While responding to these dilemmas, participants were subject to an increased cognitive load through a concurrent digit-search task. After conducting the experiment on 82 subjects, Greene et al. found that cognitive load did indeed selectively increase RT for utilitarian judgments. They observed that cognitive load increased the RT of utilitarian judgments by three quarters of a second, while there was no observable increase in the RT for non-utilitarian judgments. This confirmed the dual-process theories hypothesized asymmetry between utilitarian and non-utilitarian moral judgments.

In addition to these results, Greene et al. conducted a follow-up analysis that explored the possibility that individuals RT to problems would vary depending on their tendency to make utilitarian or non-utilitarian judgments. Greene et al. ranked participants based on how many utilitarian judgments they made in the high-conflict dilemmas. Through this analysis it was observed that high-utilitarian participants were faster in their RT than their low-utilitarian counterparts. In addition, load did not have a significant effect on the sort of judgments made by either group. These results are particularly interesting because they seem to suggest both that mental distractions don’t change the output of our judgments, and that high-utilitarian individuals are quicker to form utilitarian judgments. This quick RT may be indicative of a kind of habituation of the agent’s alignment to utilitarianism. Overall the data of this experiment seem to support Greene’s position, insofar as it shows that one kind of moral judgment (consequential/utilitarian) is affected by cognitive load while the other is not. This, along

with the 2001 and 2004 studies mentioned earlier, provide Greene's dual-process view with a strong evidential foundation on which to rest.

Now that I have considered some of the evidence that Greene provides in favour of his account, I will move on to consider some objections presented against this position. Greene's view faces many of the same concerns that were provided against Haidt's position in the last section, including the general concerns voiced by S&K about the role of reasoning in moral judgments. Greene would likely respond to many of these concerns in a similar manner as Haidt. Therefore, In order to avoid redundancy in these responses, I will consider objections that are unique to Greene's account and which focus on some of the key assumptions which support his positions.

5.3 Criticism: McGuire et al..

In my evaluation of Greene's view I am going to focus on one particular objection to his methodology raised by Jonathan McGuire et al.¹²⁶ McGuire attempts to challenge Greene's model by re-examining the data brought forward in his original 2001 paper. In particular McGuire and colleagues look to question the impersonal/personal distinction used by Greene in his earlier paper. They argue that with further analysis we will discover that only a small number of stimuli drove the effects recorded by Greene's original study, and that these stimuli aren't reasonably divided along personal/impersonal lines. In addition to this general conclusion, McGuire provides several other concerns one might have with Greene's account including "using more emotive language and references to family members or friends in their personal dilemmas, for failing to control

for cognitive processing requirements across conditions, and for ambiguously asking whether actions were ‘appropriate.’¹²⁷

After presenting these concerns McGuire et al. go on to provide an analysis which aims to disentangle particular features of the ‘personal’ category in order to test whether Greene’s results were based on the entire category, or simply a few items within the category. In order to help dissociate certain features of the category of ‘personal’ moral dilemmas, McGuire provided a series of altered versions of Greene’s original experiment. Analysis 1 recreated Greene’s original questions, excluding the non-moral control questions, and found that the distribution of answers mirrored Greene’s initial analysis. Analysis 2 provided a more substantial alteration of the original experiment, disassociating particular features of the original scenario and running it across a broader population in order to show that “the subject analysis are not generalizable to other populations of moral dilemmas, instead being driven by just a few particular dilemmas within this set of stimuli.”¹²⁸ Analysis 3 was based on the observation that scenarios that most individuals judged ‘inappropriate’ had extremely low RT scores. As such for this third analysis, McGuire and colleagues ran the experiment again, this time eliminating those scenarios which less than 5% of subjects labeled appropriate. By removing these cases RT differences between the two categories were no longer statistically significant.

McGuire concludes that a small subgroup of poorly endorsed items within the ‘personal’ category were responsible for the observed difference in RT between the two categories. McGuire argues that this should lead us to the conclusion that Greene’s dual-process interpretation of the data is false, and that some other more parsimonious interpretation ought to be adopted. McGuire extends the implications of these findings

and argues that it challenges the general credibility of Greene's dual-process approach. He concludes that these findings provide us with "no reason to assume that emotionally salient moral decisions are processed in a qualitatively different way to those dilemmas that are not emotionally salient."¹²⁹ McGuire et al. conclude that Greene's dual-processed theory bases its plausibility on an erroneous study and should be rejected.

5.4 Response to McGuire et al..

Greene's response to McGuire et al.'s criticism begins with the concession that they did indeed identify an important misstep in the original study supporting the dual process theory.¹³⁰ The focus of Greene's response then is not to try and challenge the results of McGuire et al.'s study, but to argue that they overstep the scope of the evidence by claiming that it provides grounds for refuting the dual-process theory. Greene argues that McGuire largely oversteps his conclusion by making two basic errors. The first error comes from their failure to consider other evidence in favour of the dual-process view that does not depend on the personal/impersonal distinction. The second error is that McGuire conflates the personal/impersonal distinction as it is drawn out in Greene 2001 with the dual-process view more broadly.

As we have seen, Greene presents several pieces of evidence in favour of the dual-process view following his original 2001 article. In particular, the data from the fMRI in his 2004 study, and the effects of cognitive load from his 2008 study provide alternative reasons in favour of his account that in no way rest on the results of his first study. Given that Greene has provided these additional reasons in favour of his account, and given that McGuire et al. in no way address this data, it is clearly an error to claim

that their findings provide any sort of general refutation of the dual-process account. That said, Greene recognizes that the lack of RT differences between the personal and impersonal scenarios in McGuire et al.'s study does suggest a minor problem for his view. In particular it seems to cast doubt on the fact that the utilitarian system works slower than the automatic system, as no real difference in RT's were found in the study after the confounders were removed. Greene though has made steps to address this puzzle in his previous work. Data from his 2008 study showed that those individuals who were disposed to make utilitarian judgment were hypothesized not to need to exert as much cognitive control into order to reach their conclusions. This in turn provides an explanation for these individuals lower RT's while also explaining the non-utilitarian's high RT's when forming utilitarian judgments.

In his response to McGuire's stronger conclusion, Greene provides reasons for thinking that the fate of the dual-process theory should not be tied to the personal/impersonal distinction. Recall that for Green, the personal/impersonal distinction was meant to try and capture a more fundamental distinction, namely the difference between those scenarios that engage our emotional moral system and those that engage our rational moral system. Thus for the dual process view it is irrelevant what the trigger is for distinguishing between these two systems; it is only committed to the position that these two systems are distinct. Equipped with these additional studies and an explanation for why the dual-process theory should not be conflated with the personal/impersonal distinction, Greene has grounds for rejecting the more damaging conclusions of McGuire et al.'s critique.

Now that I have shown how Greene can resist the criticism offered by McGuire et al., I can move on to consider how the plausibility of Greene's position can affect the plausibility of *psychological non-cognitivism*. For those who are interested in further considerations of the plausibility of Greene's position, I provide a set of objections articulated by Moll and Oliveria-Souza, as well as Greene's reply to these objections in Appendix A. While considerations of this critique are interesting, for the purposes of this project it will be sufficient to note that Greene is capable of responding adequately to Moll and Oliveria-Souza's criticisms. Given that Greene's position can stand up to both McGuire and Moll's critiques, we have good reason to take the implications of this view seriously. As such, I will now turn my attention to how this position affects the plausibility of *psychological non-cognitivism*.

5.5 Implications for psychological non-cognitivism

Now that we have reviewed the details of Greene's dual-process view, we can move on to assess the extent to which the truth of this position would favour *psychological non-cognitivism*. In considering Haidt's account, I argued that any theory which relies primarily on automatic and affectively charged psychological states to produce moral judgments would be a theory which provided partial support for *psychological non-cognitivism*. This support arose from the observation that affectively charged non-propositional mental states play a fundamental role in constituting our moral judgments. As we have seen, Greene's dual-process view does to a limited extent fit this mould. His theory posits that a fast acting intuitive system functions to produce our deontic moral judgments. That said, his theory diverges from Haidt's insofar as it

postulates another psychological system responsible for producing moral judgments. As we have seen, this additional system is slow, produces low level of affect in the agent, and follows utilitarian principles to arrive at its moral judgments. In order to determine to what extent Greene's theory favours *psychological non-cognitivism*, we need to determine both how prominent the utilitarian feature of Greene's position is, and the extent to which this psychological system is amenable to *psychological non-cognitivism*. In particular we need to consider whether this system functions in a fundamental propositional manner, or whether it, like the deontic system, is constituted by a non-propositional mental states.

It is hard to determine precisely how prominently the utilitarian feature of Greene's account factors into his explanation of moral judgment. Since this system is only activated in certain scenarios, the degree to which this system is active is largely determined to the sort of scenarios we encounter. As such, it is impossible to determine a priori the degree to which we will activate the utilitarian moral system in forming our moral judgments. That said, Greene did find that those individuals who are prone to utilitarian judgments (labeled 'high-utilitarians' in his 2004 and 2008 study) activate this system frequently and automatically. For these individuals at least, the utilitarian system will play a large role in the majority of their moral judgments. Thus for these individuals, and arguably even for those individuals who are not considered high-utilitarians, the utilitarian system will play an important role in determining the nature of the psychological states which constitute our moral judgments. Thus whether or not Greene's dual processing view supports or refutes *psychological non-cognitivism* will be determined largely by the extent to which the utilitarian system either favours or rejects

this position. As we will see, the question of whether or not the utilitarian system supports *psychological non-cognitivism* largely comes down to the question of whether or not this system represents our moral judgments propositionally.

In describing the features of the utilitarian system of moral judgment, Greene attributes to it several characteristics that overlap with the secondary features of cognitive mental states articulated in Chapter 3. Greene took the utilitarian system to be both slow and deliberative, involving conscious control to reach its conclusion often in the face of negatively charged emotional states. While Greene does attribute these secondary features to this system, he does not directly discuss whether the content of our utilitarian judgments should be considered propositional in nature. Recall that for a state to be propositional, it has to have communicable linguistic content that is truth sensitive. The question then becomes whether the process demarcated by the utilitarian system operates using information that is truth sensitive and functions in a manner consistent with standard propositional practices in language and logic. I argue that this system's judgments do cohere with these propositional practices. Recall that for the utilitarian, the morality of an action is determined by its overall effect on welfare. Determining whether or not a scenario will lead to a maximization of welfare requires us to be both sensitive to features of the world and to cognize these features in a way similar to other typical cognitive mental activities. While these calculations may become faster over time, the speed of these judgments is arguably attributable to a habituation of rational activity, as opposed to some non-propositional automatic response. Thus I argue the utilitarian system in Greene's account ought to be considered an example of a cognitive mental process.

Given these observations, it seems that Greene's moral view portrays moral judgments as both a product of a 'cognitive' and a 'non-cognitive' mental system. The question then becomes, if Greene has both a cognitive and non-cognitive system at work in his theory, what does this theory say about the plausibility of *psychological non-cognitivism*? This question is a difficult one, and in order to address it adequately we will have to revisit some considerations discussed earlier. In particular, we need to consider the fact that in the previous chapter we explored a strategy for having non-propositional mental states enter into propositional moral reasoning. In Haidt's system this resulted in the agent representing to themselves the non-propositional intuition of 'an aversion to x' as the propositional claim 'x is wrong'. In this case an important component of the meaningful content of the claim 'x is wrong' came from the agent's experience of the non-propositional intuition. Haidt's support for *psychological non-cognitivism* originated from the fact that the central mental state being expressed by a moral judgment was non-propositional, even when it was represented in a propositional mental state.

If Greene's theory is going to support *psychological non-cognitivism*, he will have to analogously provide a theory that requires non-propositional mental states to play a central constitutive role in our moral judgments. While the deontic/emotional side of Greene's dual-process theory likely supports *psychological non-cognitivism* in a manner similar to Haidt's SIM, the question becomes whether the utilitarian side of his theory allows for a non-propositional mental state to play a role in constituting moral judgments. As we have just observed, the Utilitarian component of the dual-process view clearly processes information in a way that is propositional. The question then is whether or not the Utilitarian system includes a non-propositional element in its content of moral

judgments. I believe that providing a definitive answer to this question is both extremely difficult and highly contentious because it rests on a central point of disagreement between the cognitivist and the non-cognitivist.

In order to show that the utilitarian judgments necessarily rest on some non-propositional component, the non-cognitivist would have to identify some essential feature of these judgments which is non-propositional. As we have seen, utilitarian judgments operate in a largely propositional fashion. Thus the only feature of the utilitarian position that the non-cognitivist could claim is non-propositional is the utilitarian's motivating premise that we ought to pursue the maximization of pleasure and the minimization of pain. The non-cognitivist could argue that the normative force of this 'ought' claim is conditional on it coming from some sort of non-propositional assertion or expression of a desire. Without this non-propositional feature, the non-cognitivist would argue, this assertion would lose some of its fundamental conceptual content, including the feeling that we ought to do it.

The cognitivist on the other hand would reject this assertion, and provide an alternative explanation wherein our commitment to utility is founded on a belief that is capable of providing the 'feeling that we ought to do it' that characterizes moral judgments. Since it is far beyond the scope of this project to determine which of these approaches is the most plausible, we will have to remain neutral on whether or not an individual's commitment to the principle of utility has to be founded on a non-propositional mental state. Since we cannot reach a determinate judgment about this point, we also cannot reach a determinate judgment on whether or not the utilitarian component of the dual-process view favours or rejects *psychological non-cognitivism*.

For the purposes of this project then, Greene's dual-process theory provides both support and potential opposition to *psychological non-cognitivism*. It provides support to this position, insofar as the emotional side of the theory postulates that non-propositional mental states play a constitutive role in moral judgments. It provides potential opposition to this position if utilitarian judgments turn out to be founded on propositional beliefs that do not incorporate non-propositional content. Now that we have considered the ways in which Greene's position can count both for and against *psychological non-cognitivism*, we can go on to the next section wherein we will examine Jorge Moll's theory and consider its implications for this position.

Chapter 6 Moll- The Event-Feature-Emotion-Complex View

6.1 Exposition

In their 2005 article *The Neurological Basis of Human Cognition*, Jorge Moll et al., provide an account of moral judgment that looks to incorporate several key neuroscientific investigations into the nature of moral cognition.¹³¹ In this article, Moll et al. look to critique several contemporary explanations of moral judgment and provide their own account which they argue best fits the neurological data. They label their theory of moral judgment the event-feature-emotion-complex view (EFEC). According to this theory, moral judgments arise out of three main features of our psychology: structural event knowledge, social perceptual functional features, and central motive or basic emotional states. Structural event knowledge is constituted largely by activity in the prefrontal cortex. These activities include structuring our observations such that they can be interpreted by other brain regions, storing ‘over-learned event sequences’ such as routine tasks, coordinating different brain regions to reach general goals, and generally performing the complex and long duration multi-staged cognitive activities necessitated by moral judgments. These activities are located primarily in the prefrontal cortex (PFC). Social perceptual and function features of moral judgments are constituted by activity in the superior temporal sulcus (STS). The STS functions to enable our interpretation of the social perceptual and functional features of our environment. The STS automatically interprets social perceptual features derived from “facial expression, gaze, prosody, body posture and gestures.”¹³² Activation of the STS is necessary for many of our moral judgments, as the cognition of social situations is often necessary in order to evoke the relevant moral sentiment.¹³³ The motivational and moral emotional features of moral

judgments are constituted largely by activity in the limbic/paralimbic and brainstem structures, as well as activity in the hypothalamus. The limbic/paralimbic structures produce basic emotional or motivational force, which is at play both in moral judgments and in other features of motivation.¹³⁴ Several of these sets of motivational structures have a strong influence on behavior due to their reciprocal connection with other brain regions such as the PFC. One of the central features of the EFEC is that it posits that a key feature of our moral judgments arise from these integrated cortical-limbic networks. These cortical-limbic networks enable us to go from noticing someone in pain (PFC/STS activity) to instantly activating limbic/paralimbic structures that trigger anxiety and attachment. This in turn leads us to consider activity which will help the person in need. Moll et al. argue that the processes outlined by the EFEC can accommodate several key features related to moral judgments.

In particular, Moll et al. focus on how their account can accommodate the emergence of moral emotions, moral values, and long-term goals. Moral emotions are those emotions which are linked to the interest or welfare of other individuals and include guilt, compassion, embarrassment, shame, pride, contempt and gratitude. In order to demonstrate how their account can explain the emergence of the moral emotions, Moll et al. focus on the example of the moral emotion of compassion. In this example, Moll et al. talks the reader through the mental processes that would take place according to the EFEC in our considerations of an older child stuck in an orphanage. According to the EFEC, in this case the PFC would provide the more complex propositional information about the situation, namely that the child's chances of adoption are low given her age. While the agent's PFC is reaching this conclusion, the STS is processing information

from the facial expression of the child, and other aspects of the social cues of the situation, such as the child's experience of helplessness. Simultaneously, the limbic/paralimbic regions are activated to produce feelings of sadness, anxiety and attachment to the child's situation. These processes combine to have us experience a feeling of sympathetic compassion toward the child and her situation. In addition to the moral emotions, the EFEC attempts to explain the emergence of broader moral values, like being a caring parent, or a good citizen. Moll et al. admit that the neural correlates of moral values are still poorly understood. That said, in neural studies of individuals expressing moral values, the same brain regions are implicated as those which the EFEC observes in normal moral judgments. Moll et al. also consider how the EFEC can explain the evaluation of the long term goals which are often involved in moral judgments. Moll et al. observant that both the activation of the anterior PFC and limbic structures are necessary in order to weigh different choices in social situations. The activation of these brain regions are predicted both by the EFEC and with other cognitive neurobiological models for reward anticipation. This view of the role of the PFC in moral judgments goes against Greene's assertion that the PFC functions to suppress emotional responses and carry out abstract reasoning. Moll et al.'s explanation of moral judgment focuses on our ability to represent and evaluate a large number of outcomes, outcomes which generate emotive force due to the links formed with the limbic systems. Thus the EFEC attempts to provide an explanation of moral judgments which integrates neurological observations from studies on the moral emotions, moral values and long term planning. Each of these features of our moral judgments rely on the three key brain regions outlined

in the EFEC: the PFC for representations, the STS for social and contextual cues, and the limbic/parlimbic region for affective and motivational responses.

6.2 Evidence

As was mentioned in the exposition of their position, Moll et al. based the EFEC on the numerous neurological studies that have been done on moral judgments. Their position looks to provide a coherent explanation of the numerous brain regions that have been shown to be activated by moral judgments. Given the vast number of studies which Moll et al. cite in favour of their position, it won't be practical to examine the experimental procedure of each. Instead, I will briefly list the findings of these studies and then explain how they support Moll's position. I will then consider more general reasons for why we might find Moll et al.'s position plausible.

Moll et al. rely primarily on evidence from fMRI studies of moral cognition to support their approach. These studies show that the anterior prefrontal cortex (aPFC), the STS, the anterior temporal lobes (aTL) and limbic structures are all implicated by moral judgments. Activation of these brain regions were observed over multiple studies.^{135 136 137 138 139 140 141 142} In addition, regions overlapping between the PFC, STS and aTL are also activated.^{143 144 145} Each of these neural observations support Moll's position by demonstrating that the three regions outlined in the EFEC play a prominent role in forming individuals moral judgments. In addition, the observed neural overlap between these brain regions supports the EFEC's assertion that moral judgments are a product of the integration of these three neural systems.

There are two additional considerations outside of this empirical evidence that arguably favour Moll et al.'s approach. First, it is able to offer a unified theory of moral judgment and second, it is able to provide a plausible explanation of the evolutionary origins of these judgments. Moll et al.'s position looks to challenge the assertion, advanced by Greene and others, that moral judgments involve multiple competing systems of moral evaluation. As we have seen the EFEC provides a more unified conception of how our moral judgments operate. Unlike Greene's position, each of the three cognitive systems of the EFEC operate together in all cases of moral judgments. The outputs of these systems are always integrated, and while different scenarios might require greater activation of one system rather than another, each system is always required in order to form a moral judgment. Moll can accommodate Greene's observation that we respond to different types of moral scenarios differently by arguing that different moral problems require a heavier reliance on certain cognitive systems implicated by the EFEC. For example, more utilitarian judgments would require additional cognitive framing, and thus would require more activity in the PFC. As such we would expect an increase in cognitive load to affect these judgments, as cognitive load is designed to target mental activity in the PFC. Moll et al.'s account is therefore shown to be able to explain the differences in these kinds of moral judgments without requiring us to posit the existence of multiple independent systems of moral cognition. While these considerations by no means settle the debate between Moll et al. and their opponents, their unified account is arguably more parsimonious than one which requires multiple competing cognitive systems. This parsimony is arguably a virtue of Moll et al.'s approach, and as such should count partially in its favour.

Moll et al.'s system also seems to gain some additional plausibility by explaining how moral judgments could emerge out of cognitive systems that play other important roles in our social and emotional mental activities. As Moll, Greene and Haidt all recognize, our pro-social moral emotions all have roots in our evolutionary history. As such moral judgments as we know them today would plausibly arise out of the integration and development of several other cognitive mechanisms designed for other tasks. The STS for example is theorized to play a role in complex social interactions, which would have played an important role in increasing the fitness of our highly social hominid ancestors. The development of the PFC in humans allowed us to do far more complex higher order cognition, which in turn allowed us to frame situations in a way distinct from our ancestors. Given these adaptations, it makes sense that our moral judgments would arise from these other newly developed features of our psychology. Moll et al.'s unified system therefore gains some plausibility from the fact that it can explain how moral judgment could emerge from our evolutionary past. These considerations, alongside the substantive amount of neuro-scientific data integrated into the EFEC provide Moll et al. with a strong evidential foundation on which to rest their position.

6.3 Criticism: Casbeer and Hynes

For the purpose of critiquing Moll et al.'s position, I will focus less on those critiques which target the mechanisms of EFEC, and more on those which target Moll et al.'s approach on the grounds of broader methodological concerns. These broader methodological criticisms will have more relevance to the discussion at hand, as they will relate to the plausibility of Greene and Haidt's positions as well. The criticism I will

consider come from observations written by William Casbeer¹⁴⁶ and Catherine Hynes¹⁴⁷ in response to a separate article written by Moll on the nature of the moral emotions. In this article, Moll et al. provide a similar explanation of the moral emotions as the one just discussed.¹⁴⁸ Casbeer and Hynes critique revolves around concerns about how to determine which sorts of judgments and emotions should count as moral as oppose to merely pro-social. Moll et al. delineate moral emotions from their 'primary' counter parts by the fact that they are other-oriented. Moral emotions then are social and other focused, while basic emotions are self-directed. As Hynes notes, this definition of what constitutes the 'moral' emotions seems to be a bit too broad, as certain cases of guilt or shame may play a very social other-focused role, but have little to do with morality. Hynes provides the example of a soccer player who misses a key shot and feels guilt and shame as a result of his failure. This feeling plays a social role as it affects his fans, and will likely have a future affect in how his team treats him (he may be fielded less often if he is determined to be an inadequate player). While such emotions may be social, they clearly are not in any relevant sense moral. Thus we need a more rigid set of criteria to delineate the social from the moral. The worry is that if we fail to delineate moral aspects of cognition properly, then we might either ignore key features of our moral judgments which aren't always present in our other social emotions, or we might overemphasize features that are relevant to social, but not to moral mental activity.

In response to this observation, each author provides their own way of delineating the domain of the 'moral'. Hynes argues that moral judgments/emotions should be defined by those behavioral imperatives which are designed to counteract selfish social behavior.¹⁴⁹ One might worry that this definition might be too broad, as it is possible that

some pragmatic means for preventing selfish behavior might not fall within the moral domain (some policy or taxation for example). In response to these sorts of concerns Casbeer argues that we will need to refer to a particular normative theory in order to demarcate the field of study.¹⁵⁰ In other words, we should appeal to certain pre-existing moral concepts like the welfare of individuals (utilitarianism) the rights of individuals (deontology) or individuals virtues (virtue theory) in order to determine which emotions/cognitive apparatuses are at play in moral judgments. One worry of this more constrained approach is that it will allow our pre-theoretic assumptions of morality to determine what we investigate, and this will in turn cause us to miss out on key observations about the mental processes involved in moral judgments. Casbeer attempts to address this concern by drawing an analogy to the study of the cognitive activity surround facial recognition. At first this study was limited to what we normally think of as 'faces' but further investigation showed that those features we wouldn't normally think of as a 'face' (ie. Two black dots and a line) activate the same regions in the brain.¹⁵¹ Thus delineating our field of investigation by some pre-theoretic notions can often lead individuals to conclusions outside those original assumptions. Thus Casbeer argues that our best means of investigating the neural correlates of moral emotions and moral judgment is to delineate our field of investigation by particular moral theories and then to continue our investigation from there. Given that Moll and others fail to delineate their studies in this manner, both Casbeer and Hynes argue that we may be failing to properly hone in on the relevant area for investigation.

6.4 Response to Casbeer and Hynes

Moll et al.'s published response to this aspect of Casbeer and Hynes objection is rather brief, and simply restates that their findings follow the standard delineation methods of cognitive neuro-science.¹⁵² Thus, in responding to Casbeer and Hynes' concern, it will be necessary to say a bit more to justify why the investigative practices of Moll, Greene and Haidt, correctly target moral phenomena. I argue that these individuals properly capture the constraints of moral judgments by designing their experiments around normal societal constraints of what is considered moral. For example, much of the behavioral data incorporated by Haidt and Greene's accounts arises out of experiments which incorporate moral thought experiments such as the footbridge scenario and the crying baby scenario in their procedures. These thought experiments are selected precisely because they are commonly recognized as moral dilemmas, both by the experimenters and the subjects involved in the experiment. Thus the psychological phenomenon these individuals are testing for are largely predetermined by what most people take to be morally important.

The data used by Moll et al. is a bit more complicated, as many of the studies they use to support their account were not originally designed to explain moral judgments. For example, much of the fMRI data Moll et al. rely on come from fMRI studies done on features of human cognition which are not explicitly concerned with moral judgments. These studies include investigations into individual's emotional reactions, their ability to pick up social cues and their ability to plan and think long term about the future. While each of these studies are not directed at moral judgments, Moll et al. adopt other considerations to help tie these non-moral cognitive process to moral judgments. They

look to accomplish this goal by coordinating this information with studies done on individuals with particular brain lesions and other disorders which effect moral judgment. In these experiments, subjects are asked questions and evaluate scenarios which are built from general social beliefs about what is moral. Thus by comparing this data with the data on more general cognitive systems, Moll et al. can make inferences about which cognitive systems are involved in moral judgments. As such the data compiled by Moll et al. is directed at our common understanding of what falls in the domain of the 'moral' because the data is structured around studies which adopt these notions. We therefore see that the fMRI data of Moll et al.'s position manages to stay close to our everyday conception of morality by relying on its use in the design of the experiments used to support his theory. This in turn shows that Casbeer and Hynes' concerns are largely addressed by the procedural methods adopted by the experiments referred to by Moll et al.'s position. More generally, this gives us good reason to think that the findings of both behavioural and neuroscientific moral psychology are properly focused on moral phenomenon, as these studies are largely built around common social notions of what is morally important.

6.5 Implications for psychological non-cognitivism

Now that we have discussed the evidence Moll et al. put forward in favour of their account, we can go on to consider the implications of their position for *psychological non-cognitivism*. As we have seen, Moll et al.'s theory provides a more integrated approach to explaining the psychological basis for moral judgments. Unlike Greene and Haidt, Moll does not draw a clear line between the rational and the intuitive features of

our moral judgments. Instead, Moll's position integrates multiple regions of the brain into a single system responsible for processing moral judgments. Moll's position also posits the existence of integrated cortical-limbic networks, which tie higher order cognitive representation to motivational states. According to this feature of his position, abstract representation of moral situations will automatically lead to the motivational and emotive force often associated with moral judgments. This feature of Moll's position opens up the possibility that propositional mental states constitute our moral judgments, and affectively charged non-propositional mental states merely get triggered by these judgments. This interpretation of Moll's position would hold that when we feel anger or empathy toward a particular situation, it can help to motivate a moral judgment, but that these motivational components have nothing to do with what *constitute* the judgment itself. This in turn seems to allow Moll's position to offer strong opposition to *psychological non-cognitivism*, insofar as it could allow for higher order propositional content to be the main kind of mental state expressed by moral judgments.

While Moll's position may at first appear to favour a cognitivist reading of moral judgments, I argue that further investigation will lead us to find that this is not the case. Using the same strategy articulated in our discussion of Greene and Haidt's positions, we can interpret Moll's position as demonstrating that the PFC functions to represent the non-propositional affective mental states in a propositional form. Thus, like Haidt, these non-propositional components will play a fundamental role in constituting our moral judgments and will be represented propositionally for the sake of communication and formal reasoning. I argue that this interpretive strategy is ultimately favoured by Moll's position, as it holds that non-propositional mental states play an important *constitutive*

role in our moral judgments. As Moll points out, there is a certain experiential feature of moral judgments which is not wholly propositional, but something that is more analogous to a non-propositional feeling. When we condemn something as wrong or praise something right we are, in the majority of cases, not simply stating a fact, we are expressing approval or condemnation of that fact. In short, we don't cognize the idea that the murder of an innocent person is wrong in the same way we cognize the fact that $2+2=4$. The experiential difference between the moral judgment and the mathematical one is that the moral judgment evokes a certain feeling about the action, while the mathematical judgment largely leaves us cold. Thus our moral judgments are not simply motivated by our emotions, but are in part constituted by them. Given these considerations, we can reach the conclusion that Moll's position provides general support for *psychological non-cognitivism*. As we have seen, this conclusion can only be reached if the general strategy advanced in Chapter 4 holds. As such, it will be prudent to take this opportunity to consider an additional objection to this strategy and examine whether or not this objection undermines the conclusions reached thus far.

One objection that might be raised against the assertion that non-propositional affective mental states play a fundamental constitutive role in moral judgments is that this view seems to contradict the belief that it is possible for us to have a genuine moral judgment without having any sort of emotion or motivational state attached to it. This is problematic because this belief is one of the central commitments of moral judgment externalism, one of the two prominent positions involved in metaethical debates concerning the relationship between motivation and moral judgments. Moral externalism is the position that a moral judgment and the motivation caused by these judgments can

be separated. In other words, the externalist holds that we can have a genuine moral judgment without having any sort of emotive or motivational response to that judgment. The central thesis of the moral externalist then seems to contradict one of the main premises of the argument we considered earlier in Chapter 4. If we can have a genuine moral judgment without any sort of motivational or emotive component, then it seems that the fundamental constitutive feature of morality has to be propositional. Thus if moral judgment externalism turns out to be true, then it seems to cast serious doubt on the central strategy I articulated in Chapter 4. This puts me in an uncomfortable position, as it is clearly beyond the scope of this project to attempt to provide any definitive judgments on the plausibility of moral judgment externalism, but without such a judgment it seems that I cannot know whether or not the central argument I have been relying on is plausible. In order to resolve this dilemma I will need to try and show how this strategy can be considered plausible regardless of the truth of moral judgment externalism. In order to see how this goal could be accomplished, we must first consider the goals of the judgment externalist.

Moral judgment externalists want to argue that the genuine features of moral judgments are not those which motivate us, or cause us to have a particular feeling, but those which we represent propositionally as 'reasons' in favour or against a position. Thus the externalist makes a substantive conceptual claim about what features ought to be included in our definition of a moral judgment. The conflict that my position generates with the judgment externalist therefore arises from what we think ought to be included in our definition of a moral judgment. The question then becomes how can we compare the plausibility of the judgment externalist's definition with one that is more

favourable to non-propositional mental states? As we saw when considering Kauppinen's objections in Chapter 1, claims about the proper limitations of our concepts are best resolved by a kind of reflective conceptual analysis, wherein competent users apply the terms in ideal conditions. While most individuals will look for this method to reach a single definition of any given term, I argue that two separate conclusions about the proper definition of 'a moral judgment' will be reached. This analysis will reach two separate conclusions concerning the proper definition of 'moral judgments' because the way we conventionally use and think about moral judgments allows for these two interpretations. These two different interpretations arise when the competent speaker focuses on different features of a moral judgment.

The first definition can be derived from the competent speaker's reflection on what it is like for them to have a moral judgment in normal conditions. This definition will align with my portrayal of Moll's position, and will state that the non-propositional feelings that accompany our moral judgments constitute a substantive part of what it is like to have these kinds of judgments. The second definition will be derived from the competent speaker reflecting on how we use moral judgments in our language and our day to day moral reasoning. This definition will be in line with the main commitments of the externalist, holding that a moral judgment stripped of its emotive or motivational feel is still a moral judgment. Thus, I argue that a reflective analysis will reach both of these conclusions because they each fulfill a different role that 'moral judgments' play in our normal discourse. The first definition looks to capture what most individuals experience when they form a moral judgment, while the second definition focuses on the propositional outcome produced by that experience. I argue that since the focus of each

of these definitions is different, an individual can adopt both definitions without contradiction. For our purposes, we can label the first definition, which includes the emotional features of standard moral judgments 'moral judgement1', and we can label the other purely propositional definition 'moral judgment2'. While it is arguable that both of these definitions capture most of what we mean by a 'moral judgment' I need to provide some reason for thinking that we ought to adopt both definitions, and not simply favour (2) over (1) as the externalists do.

I argue that we ought accept both of these definitions, because together they better capture normal discourse about our moral terms. In other words, a pluralist perspective on the definition of moral judgments is sensible because people in practice use 'moral judgment' to refer to both definitions. The fact that individuals use both these definitions is best demonstrated by otherwise seemingly contradicting statements such as 'while I rationally judge something to be wrong I don't *really* think that it is wrong.' This seemingly self-refuting assertion can be explained by the fact that an individual can judge rationally (2) that something is wrong without necessarily having the feeling that it is wrong (1). Analogously one might judge a piece of art to be good on the basis of some standard criteria, while also not really believing that it is good because it fails to generate any sort of positive aesthetic experience. I argue that differentiating between (1) and (2) enables us to explain the multiple ways in which we discuss moral judgments.

This plural view does not eliminate the possibility that, in light of additional considerations, we could limit ourselves to using only (1) or (2) as a definition of moral judgments. For example, a moral realist might claim that the feelings accompanying our moral judgments are not important because our moral judgments primary function is to

be receptive to certain moral truths, and doing so only requires definition (2). Likewise a bold non-cognitivist might argue that our moral judgments are always embedded in a non-propositional experience, and so cannot be properly captured by the propositional structure advocated by (2). That said, these more substantive claims are clearly beyond the scope of this discussion. For our purposes it is only necessary to note that overall a pluralist interpretation of the definition of 'moral judgment' is possible, and would allow for Moll's system to accommodate the truth of moral externalism by interpreting it as a position which asserts that we only ought to focus on definition (2). Thus Moll's position would be that, while the moral externalist captures some feature of our definition of a moral judgment (2), it fails to appreciate the other important features (1) which play a prominent role in our general notions of what a moral judgment entails.

Given these general considerations, we can see how the strategy I have been using can be advanced without making any substantive claims about the plausibility of moral judgment externalism. It is interesting to note that Moll's position can add further plausibility to this distinction, by providing a plausible psychological explanation for why we would have both 'cold' propositional beliefs about right/wrong and immediate emotional/motivational judgments about what is right/wrong. Recall that according to Moll, in day to day cases of moral judgments, our cortical-limbic networks function to form judgments that have both emotional force and propositional content. When reflecting on these judgments, we would be able to represent them using the PFC and abstract away from the emotive and motivational force produced by the cortical-limbic networks. This process could lead us to think of moral judgments as being strictly

propositional, and not being tied to the affective features that normally accompany these judgments.

While Moll's psychological system can explain why we might occasionally think of moral judgments as a purely propositional affair, his account clearly shows that defining all moral judgments exclusively in terms of these propositional structures produced by the PFC would be a mistake. This would be a mistake, because these propositions get their content in part from the conclusions reached by the affectively charged cortical-limbic networks. This observation vindicates the general strategy advanced in Chapter 4, as non-propositional content is shown to be a necessary part of the mental states which constitute our moral judgments.

Given these considerations, what conclusions can we reach about Moll's support for *psychological non-cognitivism*? Like Haidt, Moll allows for both cognitive and non-cognitive mental states to play a role in constituting our moral judgments. There are clearly features of moral judgments that are in some sense propositional; they have to do with the abstract slow and communicable processes of the PFC. That said, like Haidt and Greene, Moll's position has non-propositional mental states play a central role as well. Thus Moll's position, like Greene and Haidt's, provides partial support for *psychological non-cognitivism*. It is interesting to consider to what extent Moll Greene and Haidt's position support *psychological non-cognitivism*, given that each position, in its own way, allows for cognitive mental states to play a role in the formation of moral judgments. In the next sub-section of this chapter, we will explore this question in detail, and attempt to become clear on the features of *psychological non-cognitivism* these three positions support.

6.6 To What Extent Do Haidt, Greene and Moll Support Psychological Non-Cognitivism?

Now that we have reviewed each of these three positions individually, we can move on to consider what general conclusions we can reach concerning the plausibility of *psychological non-cognitivism*. In order to explore this question it will be helpful to revisit precisely what this thesis is committed to. Psychological non-cognitivism can be stated formally as follows:

(PNC) The states of mind traditionally expressed by moral judgments are not cognitive.

As we have seen, Haidt, Greene and Moll's positions each provide varying levels of support for PNC. Each view provides an explanation of moral judgments which implicates a set of non-cognitive mental states in its explanation. Each position argues that these mental states are not merely the cause of moral judgments, but in part constitute what it is for a mental state to be a moral judgment. Haidt's position for example posits that an automatic and affectively charged mental mechanism operates to produce our moral judgments. This process is not concerned with the conscious or rational contemplation of particular propositions, but instead operates automatically driven largely by social and emotional cues. These intuitions not only cause our moral judgments, they constitute them as well, providing them with their unique experiential content and motivational force. Greene's account provides similar support for the importance of non-cognitive mental states in constituting our moral judgments. For Greene the subset of 'personal' moral judgments relies on activation of particular automatic emotional responses. As in Haidt's case, these judgments are not merely

triggered by the activation of the moral emotions; these emotional responses constitute a key feature of what it is to have a moral judgment in the first place. As we have seen, Moll et al.'s theory supports a similar conclusion, implicating the activation of moral emotions and the limbic system into the standard experience of having a moral judgment.

These three theories all provide support for the notion that non-cognitive mental states are expressed by moral judgments. That said, each of these accounts also threaten the plausibility of PNC by allowing for situations where cognitive mental states play an important role in partially constituting our moral judgments. As we have seen, Haidt allows for this through his reasoned judgment link, wherein rational propositional thinking can lead us to have a new judgment. Greene's dual-process theory allows for rational-utilitarian judgments to be carried out by the propositional reasoning produced by the prefrontal cortex. In addition, Moll's position allows for propositional representation to play a role in the cortical-limbic networks which play a prominent role in our moral judgments. Each of these three positions would also be amenable to the notion that the content of moral judgments can be represented propositionally independently of the original experience of having a moral judgment. In other words, these individuals will allow for both 'moral judgment1' and 'moral judgment2' to be plausible products of the mind. Given that these three positions, and the evidence they represent, allow for both cognitive and non-cognitive mental states to play a constitutive role in our moral judgments, the question is to what extent can these three positions count in favour of PNC?

In its current form, these three judgments clearly do not support PNC. As we have just observed, each position allows for propositional mental states to play some role

in standard moral judgments. What then is the implication of these findings? Should we conclude that PNC is implausible, and as a consequence vindicate psychological cognitivism? I argue that this too would be a mistake. As we have seen, Haidt Moll and Greene each have non-cognitive mental states play a central role in our moral judgments, and as a consequence the psychological cognitivists' claim that most moral judgments are cognitive is also mistaken. Thus the first conclusion we can derive from our analysis of these three positions is that neither robust psychological cognitivism nor robust psychological non-cognitivism are vindicated by the data. Instead, we have to look for a position that is more moderate, and which allows for both types of mental states to play a role in moral judgments. To this end I argue it will be helpful to re-define PNC into a more plausible position.

In redefining PNC, it will important to try and eliminate those features of PNC which conflict with the data we have considered. The primary way in which PNC conflicts with the data is its implicit assertion that only non-cognitive mental states are expressed by moral judgments. As we have seen, Moll Haidt and Greene each recognize that cognitive mental states can play an important role in moral judgments in certain contexts. Thus in order to accommodate this data, the altered PNC has to leave open the possibility that cognitive mental states play a constitutive role in our moral judgments. That being said, the altered version of PNC has to accommodate the role non-cognitive mental states play in moral judgments. In order to balance these commitments, I argue that this altered premise should focus on the fact that non-propositional mental states play a prominent, though non-exclusive, constitutive role in standard cases of moral

judgments. This new weaker form of psychological non-cognitivism can be formalized as follows:

(PNC*) The states of mind traditionally expressed by moral judgments are in part non-cognitive.

This weaker premise acknowledges the possibility that cognitive mental states play an important role in moral judgments, while also holding that non-propositional content plays an essential constitutive role in these judgments. On the face of things, this more moderate thesis is supported by the evidence we have considered. Haidt, Greene and Moll's theories posit that a majority of moral judgments incorporate non-cognitive elements, whether they are automatic intuitive responses, moral emotions or motivational states triggered by limbic/paralimbic activity. These features of our moral judgments also help to explain the experiential character of standard moral judgments, and they help to explain why moral judgments generally provide motivation for action. It is important to note that PNC* can also accommodate the insights made earlier concerning the distinction between 'moral judgment1' and 'moral judgment 2.' While we may at first be concerned that the use of (2) contradicts PNC* by providing a set of moral judgments which does not include non-cognitive mental states, we can resolve these concerns by revising a few considerations we have already discussed. It is entirely consistent for an individual to accept that the conclusions of normal cases of moral judgment (1) involve non-cognitive mental process, while also holding that the conclusions of these judgments can be represented propositionally (2). Thus, as we saw in Haidt's chapter, it is entirely consistent for an agent to represent non-propositional content as a propositional belief. As a consequence, even if we have reason to interpret a particular moral judgment using

the definition provided by 'moral judgment 2', we can still hold that some fundamental feature of the propositions involved in (2) are non-propositional. This in turn shows us how PNC* and our earlier discussions concerning moral judgment (1) and (2) are entirely compatible.

Given these considerations, it is safe to conclude that the behavioural and neuro-psychological data provided by Haidt, Greene and Moll favour PNC*. In the next chapter we will consider the general implications of the empirical plausibility of PNC* and the concurrent implausibility of PNC. We will see that this data serves as much to sharpen the distinctions between cognitivism and non-cognitivism as it does to settle any substantive disagreement. In particular, PNC* will be shown to put pressure on the most extreme positions in the debate, and thus provide further motivation for individuals to adopt more moderate accounts.

Chapter 7- The Impact of Moral Psychology on the Moral Cognitivist/Non-Cognitivist Debate

In the last chapter we concluded that considerations from empirical moral psychology support a mediated version of psychological non-cognitivism. This mediated version holds that standard moral judgments are partially constituted by non-cognitive mental states. In this chapter we will consider the general effects of this conclusion on the cognitivist non-cognitivist debate. The first observation we should make about this conclusion is that it can be accommodated by moderate positions on either side of the debate. Insofar as a cognitivist allows for non-propositional content to play a constitutive role in moral judgments they can accommodate PNC*. Likewise a non-cognitivist can accommodate PNC* if they allow for propositional features to play some sort of role in our moral judgments. The question then becomes, do contemporary cognitivists and non-cognitivists hold these moderate positions, or do they hold stronger positions that would be threatened by PNC*? As we will see, there are individuals in both camps in the contemporary debate. One group, which I will label the 'double externalists' present a view of human cognition that directly opposes PNC* by requiring only cognitive mental states to play a constitutive role in our moral judgments. Other positions have adopted a more moderate stance and allow for both cognitive and non-cognitive mental states to play a role in constituting moral judgments. In what follows we will consider the ways in which PNC* interacts with these different positions. First we will consider how the more moderate cognitivist and non-cognitivist theories accommodate PNC* and then we will go on to consider the ways in which stronger positions might look to oppose this view.

7.1 The Effects of PNC* on Moderate Positions

PNC* is a premise which requires positions to be flexible concerning which sort of mental states play a role in moral cognition. These more moderate positions need to be able to accommodate the plausible elements of their opponent's position, and explain how both cognitive and non-cognitive mental states can play a role in our moral judgments. As we will see, both Gibbard and Smith provide positions that satisfy these criteria. Smith provides a cognitivist account which can accommodate non-cognitive mental states by making them part of what substantiates the belief states that constitute moral judgments. Gibbard's non-cognitivist account on the other hand looks to accommodate a form of predication, n-predication, which is able to explain the seemingly cognitive mental states that play a role in standard cases of moral judgment. In order to better appreciate how each of these positions can accommodate PNC*, I will revisit each account in further detail.

As we observed in chapter 2, Smith provides a unique explanation of moral judgments which attempts to synthesize the insights of individuals from both sides of the cognitivist non-cognitivist debate. Smith is favourable to Hume's sentimentalism and desire based moral psychology, and as a consequence his position is comfortable with the notion that non-cognitive mental states would play a role in the formation of our moral judgments. That said, Smith is a cognitivist and is committed to the claim that the belief 'what one would do if one were fully rational' plays a central role in forming our moral judgments. Recall that Smith chose this belief because it has an important relationship to our desires, which Smith argues is necessary in order to account for the motivational force of our moral judgments. This belief's connection to our desires also arguably

connects it to the non-cognitive mental states associated with these desires, such as emotional states like fear and joy. Thus for Smith, it is natural for non-cognitive mental states to play a role in what it is to have a standard moral judgment.

While this explanation arguably shows why non-cognitive mental states would play a role in *influencing* our moral judgments for Smith, one might worry that it does not do enough to show that these states partially *constitute* our moral judgments as well. It could be argued that the proper interpretation of Smith's position would be to view moral judgments as constituted by the belief 'that which I would do if I were fully rational' and to in turn view non-cognitive mental elements as mere consequences of entertaining this belief. I would argue that this interpretation of Smith is problematic as it fails to properly understand Smith's position as it arises in *The Moral Problem*. Recall that the reason Smith believes the belief he selects can solve the *moral problem* is because it bears the right kind of relationship to our desires. The important motivational force provided by a desire is not produced by a propositional mental state, but by something akin to a non-propositional expression of preference. Thus Smith already allows for the fact that our moral beliefs motivate us and are important to us because they include the non-cognitive mental states which constitute desires. Given this consideration, and given Smith's general affiliation with Humean sentimentalism, we would expect that his position would support the notion that moral judgments are partially constituted by non-cognitive mental states. Smith's moral judgments *include* non-cognitive mental states, because these states are part of what makes the belief 'what I would do if I were fully rational' motivating in the first place. Without including non-cognitive mental states as *part* of what it is to have a moral judgment, the belief about 'what I would do if I were fully rational' would have

no motivational force and would thus fail to solve the *moral problem*. For these reasons I argue Smith's position would promote a conception of moral judgment that includes both cognitive and non-cognitive mental states, and as a consequence can accommodate PNC*.

Gibbard's non-cognitivist position also provides room for both cognitive and non-cognitive mental states in the production of our moral judgments. As we have seen, Gibbard's position revolves around considerations of plans for action. The non-cognitive element of Gibbard's position comes from the observation that we chose to follow these plans of action as a consequence of an expressed affinity toward them. While these plans are pursued on the basis of our non-cognitive expressions of preference, Gibbard allows for pseudo-propositional entities to enter into our moral discourse. In part, this propositional component enters into our plans when these plans refer to features in the world. In this regard, our moral judgments are propositional because they make reference to activities in the world. Gibbard also looks to accommodate propositions at a more fundamental level of moral discourse by introducing the concept of n-predication. While n-predication isn't predication in the robust sense, it functions in a similar manner. As we have seen, n-predication predicates features of a particular 'normative-factual word', which is itself determined by our expressed preference for one set of norms rather than another. N-predication then is shown to be non-cognitive because it relies on our expressed preference of a particular normative-factual world. While n-predication is non-cognitive, it functions in a way extremely similar to normal predication. It is therefore not unreasonable to assume that n-predication would rely on the same cognitive mental activity as regular predication. Insofar as Gibbard can provide an adequate explanation

for why the cognitive mental states involved in n-predication shouldn't be considered cognitive in the semantic sense, Gibbard's account can accommodate the idea that cognitive mental activity plays a role in moral judgments without jeopardizing his commitment to non-cognitivism. Gibbard's position is thus able to accommodate the possibility left open by PNC*, that cognitive mental states play a role in our moral judgments.

Now that I have considered the ways in which more moderate contemporary positions can accommodate PNC*, I will move on to address those positions that are threatened by the empirical plausibility of this position. In particular I will focus on a group which I have labelled the 'double externalists' who argue that non-cognitive mental states cannot play a role in constituting our moral judgments. After considering this position I will consider some objections its proponents might put forward against the plausibility of PNC*. I will respond to these objections, and then conclude the chapter by considering the broader impact the plausibility of PNC* has on the contemporary debate.

7.2 Double Externalism and its Opposition to PNC*

There are a group of individuals within the contemporary metaethical debate who I argue take a very strong stance on the nature of the mental states involved in our moral judgments. These individuals, who I have labelled 'double externalists'¹⁵³, hold that only cognitive mental states can play a role in constituting our moral judgments. The reason I have labelled this position 'double externalism' is because its proponents are both moral judgment externalists and moral reason externalists. Double externalists argue both that moral judgments do not entail any motivational state and that our judging a moral

proposition to be true does not affect the reasons that compose these judgments. Thus, according to this position an individual can form a moral judgment without being motivated to act in a particular manner. In addition, the agents own interests play no role in determining the truth of the proper content of these judgments. Such strong externalism allows for the individual to exclude any and all non-cognitive forces that might be playing a role in moral judgments. It does so by excluding the non-propositional mental states that normally accompany motivation, as well as any non-cognitive mental states that might cause us to advance our own non-referential reasons as the content of our moral judgments. The commitments of this position thus exclude, by definition, any non-cognitive mental states playing a role in our moral judgments. Double externalists would thus object to the PNC* by arguing that a proper analysis of moral judgments reveal that it is impossible for non-cognitive mental states to play a constitutive role in these judgments.

A possible example of this position can be found in a particular interpretation of the moral theory advanced by Derek Parfit.¹⁵⁴ Parfit's theory of morality conceptualizes moral judgments as judgments first and foremost concerned with reasons. For Parfit, reasons are the basic units of normative justification for our actions.¹⁵⁵ This definition of reasons partially parallels our colloquial use of the term. When we want to explain why we did something, or why we ought to do something, we provide reasons for action. While the term is elevated for the purposes of philosophical discussion, Parfit's use of reason stays fairly close to its use in justifying everyday behaviors. Moral judgments for Parfit are concerned with one set of these reasons. According to Parfit, when we make a moral judgment we are essentially attempting to make reference to the reasons that apply

to our situation.¹⁵⁶ This explanation of moral judgments in turn helps to explain Parfit's potential commitment to double externalism. Our recognition that there is a reason for an action does not necessitate that this reason comes from our own subjective interests, nor does it necessitate any emotive or motivational response to this reason. This in turn seems to disqualify non-cognitive mental states from constituting part of our moral judgments. If moral judgments are defined as propositional mental states which refer to reasons, and reasons by definition exclude non-propositional content, then non-cognitive mental states have no role to play in moral judgments. Thus, Parfit would recognize that while emotions may influence our moral judgments, they can never be properly considered a constitutive feature of these judgments. This interpretation of Parfit's position is thus fully opposed to the truth of PNC*, as non-cognitive mental states cannot play a constitutive role in any of our moral judgments.

As I have argued in the last chapter, PNC* is a premise that is generally supported by the empirical data provided in this account. Thus, in accord with our conclusions to chapter 1, any position which commits itself to a substantive claim about human psychology which contradicts this premise will be called into serious question. As we have seen, double externalists clearly commit themselves to claims about human psychology which contradict PNC*, and as a consequence should be rejected. In order to retain their plausibility in the face of this objection, the double externalist could provide a number of arguments in order to call PNC* into question. Before we can draw any substantive conclusions about the effects of PNC* on this position, we will need to address these critiques.

7.3 Objections to PNC*

As we have seen, the plausibility of 'double externalism' is called into serious question by the empirical plausibility of PNC*. In order to counteract this objection, proponents of this position need to provide reasons for thinking that PNC* is mistaken. Since the proponents of double externalism are clearly not concerned with providing contradicting empirical data, their objections will have to focus on challenging one of the conceptual premises which PNC* relies on. One way to do so will involve revisiting the distinction from the last chapter between 'moral judgment 1' and 'moral judgment 2'. An individual looking to defend double externalism could argue that 'moral judgment 1' should be discarded and that we should focus primarily on 'moral judgment 2'. In addition, the double externalist would argue that we ought to adopt a definition of 'moral judgment 2' which only takes propositional content as its focus. If the double externalist can sufficiently motivate this position then they will have a strong case for confronting the conclusions of PNC*, as moral judgments will now, by definition, only include propositional belief states. In order to advance this conclusion, the double externalist will have to argue that the motivational or emotional components of our day to day moral judgments should be ignored. The key task for the double externalist then is to provide some reason for asserting that the propositional features are the only features worth being considered in our definition of a moral judgment.

One way the double externalist could motivate the claim that we should adopt a purely propositional definition of moral judgments is by claiming that such a definition would better track what morality *is really about*. This approach would argue that the primary function of a moral judgment is to put us into contact with the truth-makers of

our moral claims. A proponent of my interpretation of Parfit's position for example, would argue that we should limit our definition of moral judgments to a set of propositional claims that are capable of referring to object-given reasons. Parfit's objection thereby originates from a claim about the proper function of our moral language. If the only important feature of a moral judgment is that it puts us in contact with moral truth, then any non-cognitive feature which cannot perform this function should be ignored.⁹ If Parfit is successful in demonstrating that the proper definition of moral judgments is such that it can only be satisfied by cognitive mental states, then the psychological investigation is over before it has begun, since any psychological data which purported to implicate non-cognitive mental states in moral judgments could be rejected on the basis that it fails to cohere with the object of investigation. Thus a great deal is at stake in whether or not the double externalist can provide adequate reasons to limit the definition of morality to a purely propositional form of moral judgment².

As we observed in the first chapter, discussions concerning the proper constraints of our concepts need to be resolved using traditional methods of conceptual analysis. Kaupinen argued then when we attempt to define a term we should consider how a competent agent in ideal conditions would define it. The question then is whether or not a competent agent in ideal circumstances would restrict their definition of moral judgment to fit Parfit's constraints, or whether they would adopt the broader definition captured by

⁹ It is interesting to consider how this objection draws together the two theses van Roojen articulates to separate the cognitive from the non-cognitive sides of the debate. If the double externalist can win a strong victory in the domain of *semantic non-cognitivism* then this might affect the plausibility of *psychological non-cognitivism* as well, insofar as the domain of what is considered a 'moral judgment' is determined in part by the semantic constraints of this term. That said, there is still good reason to suppose that even a total victory in the first debate will leave the second debate underdetermined.

PNC*. This question in turn requires us to revisit considerations made in the last chapter concerning what features should be considered relevant in our definition of moral judgments. Empirical psychological investigation has shown us that non-cognitive mental states play an important role in our day to day moral judgments. As a consequence, most individuals understanding of what a moral judgment entails will include this non-cognitive element. Thus the majority of people will identify some sort of emotive or motivational change to be part of what it is to have a moral judgment. The question then is whether a fully rational agent in ideal circumstances would also agree to this conclusion.

The answer to this question requires us to revisit the reasons I articulated earlier for distinguishing between moral judgment₁ and moral judgment₂. As we have seen, the way in which a competent agent in ideal circumstances will define a moral judgment will largely rely on what feature of the phenomena of a moral judgment they are trying to define. If they are attempting to explain what they normally experience when they have a moral judgment, they will admit a level of non-cognitive activity, for if they are a psychologically stable individual, they will have some non-cognitive responses to judgments in morally relevant scenarios. In line with our previous analysis, we should recognize that an agent conducting this analysis under ideal conditions will also recognize that part of what it is to have a moral judgment can be captured in a propositional form. Thus one can adopt the belief 'I take x to be wrong' without, in entertaining that thought, experiencing an emotion or non-cognitive mental state. This analysis clearly shows that an agent would reach two separate definitions for what a moral judgment entails depending on whether they were focusing on the experiential

features of having a moral judgment, or the propositional beliefs that are built from these experiences. In order to block this conclusion, the double externalist would have to provide additional reasons for thinking that we ought to only adopt the propositional interpretation of our moral judgments. That said, the only reason they could provide in favour of this position is that it advances their own theoretical commitments. Since a robust analysis of what a moral judgment entails reveals both moral judgment¹ and moral judgment² as plausible definitions of our moral judgments, the double externalist cannot, on the basis of an ideal semantic analysis, claim that we ought to restrict our definition of moral judgments to propositional mental states.

I argue that these considerations show that the double externalist would not be able to limit the definition of moral judgments to propositional mental states using considerations of the descriptive character of standard moral judgments. That said a proponent of Parfit's position could attempt to limit the definition to propositional features by appealing to 'ideal-epistemic' considerations. Parfit, or any other moral realist for that matter, could argue that we ought to restrict the domain of moral judgments to those mental features which could, at least in principle, make reference to the truth makers of our moral terms. We ought to restrict our definitions in this manner because, insofar as the moral realist is right, our moral judgments are designed to track moral truths. Thus, our definition of moral judgments should be restricted to cognitive mental states, as these are the only states capable of referring to moral truth. Parfit could therefore concede that our standard conception of a moral judgment include non-cognitive mental states, while also arguing that we should ignore these mental states because they are incapable of accomplishing the epistemic goal of our moral judgments.

Thus the 'ideal-epistemic' objection rejects PFC* on the grounds that it includes features which inhibit the truth-referential role of our moral judgments.

These considerations allow Parfit to argue for a reclassification of what ought to count as a moral judgment. It is important to note that this reclassification does not look to challenge any of the empirical observations captured by PNC*. It would still be true that individuals employ non-cognitive mental states in standard cases of moral judgments. That said, this position would have us reclassify this data in order to exclude non-cognitive processes from being considered a constitutive feature of our moral judgments. It should be noted that this strategy is a tenuous one, as it not only relies on a victory for the cognitivist in the more general debates concerning moral realism, it also requires us to reclassify our judgments in a way that is not true to human psychology or to standard human experience. If moral judgments, as the data suggests, really are tied to affective non-propositional experiences in normal cases, then the most adequate theory should be one which recognizes this fact. It may be, at the end of the day, that the realist has reasons to mitigate this feature of moral judgments in order to allow moral judgments to better serve their truth-referential purpose, but such considerations should not distract us from the fact that when most people form a moral judgment they are doing something more than simply entertaining another proposition. Moral judgments have a particular experiential character, and this unique experiential character is best explained by the inclusion of non-propositional affective states into the definition of what it is to have a moral judgment. While double externalists might argue that the only important feature of moral judgments are their truth referential features, this move is problematic because it is based on the faulty assertion that the proper definition of a word can be determined by

how it looks to be used by a particular theory. While the double externalist wants to limit the definition of moral judgment until it is merely truth referential, it is an open question whether or not this concept, in our psychology and in daily use, incorporates more than this referential feature. Indeed our analysis thus far has shown that there are good reasons to think that moral judgments are not simply concerned with propositional referential activity, but function to incorporate non-propositional affective components as well. Thus I take PNC* to be a plausible commitment even in the face of this objection. If moral realism is true, we might have epistemic reasons to favour the truth-referential features of our moral judgments over the non-propositional ones, but this doesn't take away from the psychological observation that non-propositional mental states play a role in these judgments as well. Thus the double externalist's attempt to refute PNC* is unsuccessful, as even if moral realism turns out to be true, it doesn't eliminate the prominent role non-cognitive mental states play in our moral judgments.

Our discussion thus far has shown that the double externalist's position is strongly challenged by the empirical plausibility of PNC*. The double externalist may, for epistemic reasons, ignore the psychological realities captured by PNC*, but this move does not drastically change the plausibility of their position. If double externalism is going to remain plausible, it has to allow for the possibility that non-cognitive mental states play a constitutive role in at least some of our moral judgments. If it cannot find a way to accommodate this position, then in the face of the evidence in favour of PNC*, double externalism will be found to be, at least in this regard, an implausible position.

7.4 General Implications of PNC*

The majority of this chapter has been focused on working out the implications of PNC* for contemporary positions in the moral cognitivism/non-cognitivism debate. I have argued that more moderate positions on both sides of this debate can accommodate this conclusion, while more extreme views conflict with it and are thereby called into question. As we have seen, 'double externalism' can attempt to mitigate the implications of PNC* by providing epistemic reasons for limiting the domain of moral judgments to cognitive states, but such a response does not mitigate the descriptive plausibility of PNC*.

The plausibility of PNC* can be seen as adding additional pressure for individuals to incorporate the plausible features of their opponents positions in the cognitivist/non-cognitivist debate. PNC* shows us that both cognitive and non-cognitive mental states have an important role to play in our moral judgments. While PNC* does not radically alter the landscape of the contemporary debate, it does provide us with additional reasons for favouring moderate positions like Gibbard and Smith's over more extreme positions like double externalism. This conclusion in turn demonstrates how findings from moral psychology can effect positions in the domain of metaethics. By clearly articulating the psychological commitments of each position and comparing these commitments to the findings of empirical moral psychology, we can reach new and important conclusions about the plausibility of these positions. While such conclusions will always be limited by the data available and our ability to translate the metaethical positions into a psychological format, even modest conclusions like PNC* can have an important impact on framing and directing future debates.

Conclusion

In this work I have attempted to show how considerations from empirical moral psychology can affect the cognitivist non-cognitivist debate. I began this analysis by addressing some general concerns surrounding the use of empirical data in investigating philosophical problems. In attempting to address these concerns I came to the conclusion that empirical data can be applied to a philosophical position when this position commits itself to a particular empirically testable feature of the world. This in turn led to the conclusion that in order to properly apply the findings of empirical moral psychology to the cognitivist non-cognitivist debate, I needed to find some feature of this debate which was concerned with facts about human psychology. After analyzing the commitments of cognitivists and non-cognitivists, I discovered one such position, psychological non-cognitivism, which commits non-cognitivists to the position that moral judgments express non-cognitive mental states. After establishing a rough set of desiderata for distinguishing cognitive from non-cognitive mental states, I set out to test PNC empirically. I chose to test the empirical plausibility of this premise by examining its compatibility with the data provided by Haidt, Greene and Moll. My consideration of these three individuals' positions revealed that psychological non-cognitivism was too strong of a claim to be supported by the data. This led me to adopt a more moderate form of this position, PNC*, which held that non-cognitive mental states constituted a central component of standard moral judgments. This weaker premise was able to accommodate Haidt, Greene, and Moll's observation that both propositional and non-propositional elements play a role in constituting our moral judgments. While PNC* was more moderate than its unaltered counterpart, it was still shown to provide strong opposition to

more extreme metaethical views such as double externalism. Thus, by supporting the plausibility of PNC*, empirical moral psychology directly affected certain features of the moral cognitivist/non-cognitivist debate.

As I stated at the beginning of this project, my interest in examining the effects of empirical moral psychology on the cognitivist non-cognitivist debate did not arise simply from my interest in this debate, but in a general curiosity in how considerations from empirical moral psychology might affect metaethical discussions. Thus my analysis not only looked to address the cognitivist non-cognitivist debate, but it also looked to provide an example of how moral psychology might come to effect discussions happening in the field of metaethics. To this end it will be helpful to look past the particular conclusion of PNC*, and to focus on the general conceptual moves that allowed us to apply data from empirical moral psychology to a traditionally metaethical debate.

The analysis carried out in this work demonstrates that there are two general ways in which empirical moral psychology can affect metaethical debates. The first method is direct; it looks to determine whether the metaethical view in question commits itself to any empirically testable claims about human psychology, and then tests these claims using the appropriate psychological evidence. For some debates this process is straight forward, as the psychological commitments of the metaethical position are easy to identify. For example, the debate between the moral judgment externalist and the moral judgment internalist clearly has to do with a particular psychological feature of our moral judgments, namely whether or not they necessarily entail a motivating psychological state. While some debates are relatively easy to test empirically, other debates in metaethics are less clearly relatable to empirical claims. For example, questions about

the epistemic status of our moral intuitions might come down to some commitment about the nature of the psychological states that produce our moral intuitions, and their relationship with certain facts in the world. That said translating this debate in this manner requires us to make several controversial claims about both the nature of our intuitions and the role they ought to play in our conceptual analysis. Thus some metaethical positions might only be amenable to psychological data once hard conceptual battles have been won. Whether its application is complex or straight forward, this first method provides a clear manner in which the empirical findings of moral psychology can apply to metaethical debates. While I have focused primarily on this method of applying the data in this work, there is a more subtle and potentially profound way for considerations from moral psychology to affect metaethical debates.

This more subtle method forces us to clarify and challenge concepts that were useful in the context of metaethical debates but which fail to have any obvious correlation in the psychological domain. This approach looks to use empirical moral psychology to provide us with a vantage point to critique the usefulness or plausibility of some of our metaethical concepts. For example, in my analysis I was forced to provide an explanation of what ‘cognitive’ and ‘non-cognitive’ translated to in the psychological context. While I did my best to explain how these metaethical terms could be translated into the psychological domain, some may argue that the difficulty of this task indicates something about the plausibility of these concepts. Perhaps the distinction suggested by cognitivism and non-cognitivism simply is not present in human psychology, and if this is the case then perhaps we should consider either abandoning or revising these terms in order to make them cohere with an empirically informed understanding of the nature of

human cognition. While this is an exciting possibility, such a strong position can hardly be adequately defended or articulated here. For the purpose of this project, it will have been enough to show that empirical moral psychology can directly affect metaethical debates. It will be up to future investigations to determine whether this new and bolder approach can more profoundly affect the landscape of the metaethical debate.

Appendix A- Additional Considerations for Haidt and Greene

Pizarro and Bloom Critique of Haidt

In their paper *The Intelligence of the Moral Intuitions: Comment on Haidt (2001)*,¹⁵⁷ Pizarro and Bloom argue that while they agreed with many of Haidt's conclusions, they believed that the SIM places too little emphasis on the role of reason in shaping our moral intuitions. They agree with Haidt that people possess intuitively given first principals, but they argue that deliberative reason does far more to determine our moral judgments than Haidt's position would have us believe. Pizarro and Bloom focus on two ways that reasoning can affect our moral judgments, the first is by framing the situation in a way to trigger certain intuitions, and the second is by exerting direct control and altering these intuitions intentionally.

Pizarro and Bloom point out that reasoning can influence our intuitions by altering the way we evaluate a situation. They refer to a number of studies that have demonstrated the importance of cognitive appraisals in changing quick involuntary responses.¹⁵⁸ These studies provide evidence that suggest that if an individual has certain beliefs about a situation, they are far more likely to trigger particular automatic responses. For example, Dandoy and Goldstein conducting an experiment where individuals viewed films where people suffered bodily harm, and those subjects who were told to take an 'objective perspective' toward the film showed less psychological distress than those who were not primed in this manner.¹⁵⁹ Individuals often change their intuitive responses when they adopt new beliefs about the situation. For example, cases where individuals take another person's perspective, feelings of anger can often turn to

sympathy. This is consistent with the observation that individuals can instantly change their emotional disposition in response to new information. For example, when an individual is told that their friend is running late a meeting not because they are neglectful, but because something tragic has occurred in their life, their feelings instantly shift.¹⁶⁰ Experiments have also demonstrated that by getting people to think about the other person's perspective causes them to become more empathetic.¹⁶¹ Thus reason's ability to shift our focus and develop new beliefs plays a large role on the outputs of our intuitive moral reactions.

The second general way that rationality can affect our moral intuitions is by altering them directly. Pizzaro and Bloom note the growing literature that shows that we have second-order control over many of our emotional reactions and automatic judgments.¹⁶² There are a number of ways that our reason can control the more automatic processes of our nature. The first is by using our rational judgment to place ourselves in environments that will lead to change in our moral intuitions. Pizzaro and Bloom point to a number of 'common sense' examples. For instance, individuals who were exposed to positive African American Exemplars either through an experiment¹⁶³ or through an African American professor teaching a course on racism¹⁶⁴ greatly reduced their implicit negative attitudes toward African Americans. It is commonplace for us to focus on appraising a situation in a particular manner in order to evoke sympathy if our friends or family are involved. By focusing on particular features of the scenario, we are able to influence which intuitions will be triggered in that scenario. Thus Pizzaro and Bloom conclude that deliberative rationality can influence our intuitions by placing us in scenarios that will inculcate the attitudes we want, and discourage the attitudes we do not.

Through these two mechanisms, cognitive appraisal and direct cognitive alteration our intuitions, Pizzaro and Bloom argue that rationality can play an important role in shaping our emotions. While they recognize that none of these features directly contradict Haidt's model, they argue that it casts doubt on Haidt's emphasis on the non-rational features of our moral judgments. In other words, Pizzaro and Bloom take rational deliberation to be a far more important feature of our daily moral judgment than Haidt's model will allow.

Haidt's response to Pizzaro and Bloom

Haidt's response to Pizzaro and Bloom is on the whole more moderate than his response to S&K.¹⁶⁵ Unlike S&K Haidt does not take Pizzaro and Bloom's critiques to be based on an incomplete reading of his position. Instead he takes their challenge head on, attempting to provide further justification for his belief that rational reflective deliberation plays a minor role in the majority of individual's moral judgments. First Haidt echoes Pizzaro and Bloom's observation that individuals often automatically change their mind when new information about the situation comes to light. If someone was running late to a meeting because they had just suffered a loss in the family, the realization of this fact really would instantly change you're feeling about it. Haidt claims that his position is designed in part to explain why individuals come to experience new cognitive appraisals in these sorts of situations. While Bloom and Pizzaro seem to think that these new appraisals happen as a consequence of private rational deliberation, Haidt argues that real psychological systems at work are much faster and much more subtle. He cites the evidence mentioned earlier which established that individuals are particularly bad at judging both sides of a situation in an unbiased fashion.^{166 167} These observations

lead Haidt to conclude that it is not a conscious rational deliberation that leads individuals to change their attitudes when new information comes into play, but a more automatic intuitive system directing these behaviours. In addition the fact that we can rationally frame a situation in a way that will trigger one intuition rather than another does not mean it happens in the majority of cases. As Pizzaro and Bloom recognized, the fact that cognitive framing changes our intuition does not threaten Haidt's position, as it simply redirects our intuitions to different elements of the scenario. This sort of cognitive framing is difficult, and it is difficult precisely because it is not a natural feature of our moral judgments. As such Haidt argues that while we can frame things to change our moral intuitions, this is neither a necessary nor a common feature of our moral judgments. Haidt responds to Pizzaro and Bloom's second point in a similar fashion by arguing that our rational second-order judgments rarely are the cause of practices that change our intuitions. Haidt critiques Pizzaro and Bloom's own example. He argues that it is far more likely that individuals will join a class on racism, or actively seek out positive role models from a particular group, in order to feel "the joy of morally meshing with other people who are outraged by racism."¹⁶⁸ It is this pro-social desire to share in and propagate a particular moral position that in the vast majority of cases, motivate these actions, not some detached rational deliberation. Thus Haidt concludes that Pizzaro and Bloom's criticisms do not cast serious doubt on his position.

Moll's Critique of Greene

Jorge Moll and Ricardo de Oliveira-Souza focus their criticism of Greene's position on the fact that some of the key information Greene provides in support of his

position better supports a different conclusion.¹⁶⁹ In this paper Moll and Oliveria-Souza argue that Koeings recent research concerning the effects of selective damage to the Ventromedial Prefrontal Cortex (VMPFC) does not, despite Greene's claim, support the dual-process theory.¹⁷⁰ In addition, they argue that a more parsimonious explanation of the data presented by Koeing is that damage to the VMPFC causes the reduction of certain pro-social emotions.

In his study Koeings found that individuals who suffered damage to their VMPFC respond in a distinctive manner when confronted with moral dilemmas. When these individuals were confronted with 'difficult' moral dilemmas they were far more likely than normal subjects to choose the 'utilitarian' option. This is despite the fact that choosing the 'utilitarian' option resulted in highly emotionally aversive experiences for normal subjects. One suggested explanation for this behaviour is that VMPFC patients failed to experience the negative emotional states associated with these options. These considerations are derived from a number of studies analyzing the role of the VMPFC in moral judgment. For instance, Damasio et al. have shown that individuals with damage to this region of their brain maintain their ability to continue to perform formal moral reasoning, but suffer severe impairments in their daily social interactions, often leading to inconsistent and immoral behaviour.¹⁷¹ In addition it has been observed that early damage to the VMPFC leads to a severe impairment in both moral behaviour and moral reasoning. Moll observes that many individuals, including Greene, want to interpret this data to indicate that damage to the VMPFC leads to an emotional blunting which then makes it more likely for individuals to make utilitarian choices in high conflict scenarios. Moll argues that this interpretation is ruled out by another study done by Koenigs et al..

In this study, individuals play a Ultimatum game where individuals can choose to accept an unfair but financially beneficial offer (the economically 'rational' choice) or choose to reject the offer and punish the other player (the 'emotional' choice).¹⁷² VMPFC patients were found to reject the offer more often than other participants. Moll argues that these additional findings reject Greene et al.'s original interpretation because they show that damage to the VMPFC can in fact make individuals more prone to emotional decisions. Moll argues that in order to accommodate this data Greene would have to accomplish two tasks. First, Greene et al.'s position would have to show both that VMPFC damage lead to more utilitarian judgments (presumably because it reduces emotional responses) and that damage to areas responsible for moral reasoning, such as DLPFC and lateral FPC, lead to an increase in emotionally based moral judgments. Second, Greene et al. would have to explain how their view could still be plausible as the regions most robustly activated by utilitarian choices were damaged in most patients in Koenigs study. As such the dual-process view does not adequately accommodate Koenigs data, as it makes false predictions about the brain regions implicated by utilitarian judgments, and it fails to explain the more emotional responses made by VMPFC patients in Koenigs later study.

Greene's response to Moll

Greene provides a concise and direct response to Moll's objection.¹⁷³ To Moll's first point, Greene argues that while being able to provide a double dissociation between VMPFC damaged utilitarian patients and anti-utilitarian DLPFC damaged patients would provide excellent evidence for his theory, it does not mean that a single dissociation provides no evidence in its favour. Greene responds to Moll's second point by showing

that the regions usually associated with utilitarian reasoning that were missing in Koenigs' VMPPFC patients, are not a necessary feature of this reasoning. Greene et al., originally observed that these regions were likely involved in controlling countervailing emotional responses. If VMPPFC patients lacked these emotions in the first place, then there would be no need for control. These individuals will be capable of utilitarian judgments as long as they can still carry out utilitarian reasoning, and this is hypothesized to take place in the more posterior regions of the DLPFC, regions that were not damaged in Koenigs participants. Given these considerations, Greene argues that his view remains plausible even in light of Koenigs observations. Indeed, Greene argues that these studies support his position, as "Only the dual-process theory specifically predicts that VMPPFC patients will be unusually utilitarian across a range of difficult cases, and the inclusion of brain-damaged control patients demonstrates the anatomical selectivity of this effect."¹⁷⁴

Notes

¹ Kauppinen, Antti. "The Rise and Fall of Experimental Philosophy." *Philosophical Explorations*, 2007: 95-118.

² Ibid.

³ Ibid 100

⁴ Ibid

⁵ Ibid

⁶ Knobe, Joshua, and Shaun Nichols. "An Experimental Philosophy Manifesto." In *Experimental Philosophy*, by Shaun Nichols, 3-17. New York: Oxford University Press, 2008.

-
- ⁷ Knobe, Joshua, and Shaun Nichols. "An Experimental Philosophy Manifesto." In *Experimental Philosophy*, by Shaun Nichols, 3-17. New York: Oxford University Press, 2008.
- ⁸ Ibid
- ⁹ McGrath, Matthew. "Propositions." *Stanford Encyclopedia of Philosophy*. December 19, 2005. <http://plato.stanford.edu/entries/propositions/> (accessed July 20, 2012).
- ¹⁰ Joyce, Richard. *The Myth of Morality*. New York: Cambridge University Press, 2001.
- ¹¹ Ayer, A.J. *Language, Truth and Logic*. Oxford: Oxford University Press, 1936.
- ¹² Stevenson, C.L. *Ethics and Language*. New Haven: Harvard University Press, 1944.
- ¹³ Gibbard, Allan. *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. New York: Oxford University Press, 1990.
- ¹⁴ Gibbard, Allan. *Reconciling Our Aims*. New York: Oxford University Press, 2008.
- ¹⁵ Ibid. 29
- ¹⁶ Ibid, 16
- ¹⁷ Ibid.
- ¹⁸ Ibid
- ¹⁹ Gibbard, Allan. *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. New York: Oxford University Press, 1990. 95
- ²⁰ Ibid
- ²¹ Blackburn, Simon. "Critical Notice of Allan Gibbard's *Wise Choices, Apt Feelings*." *Ethics*, 1992.
- ²² Smith, Michael. *The Moral Problem*. New York: Blackwell Publishing, 1994. 7
- ²³ Ibid 12
- ²⁴ Ibid 152
- ²⁵ Ibid 182
- ²⁶ Van Roojen, Mark. "Moral Cognitivism vs. Non-Cognitivism." *Stanford Encyclopedia of Philosophy*. 2009. <http://plato.stanford.edu/entries/moral-cognitivism/> (accessed July 20, 2012).
- ²⁷ Ibid.
- ²⁸ Ibid.
- ²⁹ Johnson, Gregory. "Theories of Emotion." *Internet Encyclopedia of Philosophy*. June 9, 2009. <http://www.iep.utm.edu/emotion/> (accessed July 20, 2012).

³⁰ Ibid

³¹ Ibid

³² Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6, 169–200.

³³ Haidt, Jonathan. "The Emotional Dog and its Rational Tail: A Social Intuitionist Approach to Moral Judgments." *Psychological Review*, 2001: 814-834.

³⁴ Graph from: Haidt, Jonathan. "The Emotional Dog and its Rational Tail: A Social Intuitionist Approach to Moral Judgments." *Psychological Review*, 2001:815

³⁵ Ibid 814-834

³⁶ Nisbett, R.E., & Wilson, T.D. "Telling more than we know: Verbal reports on mental processes." *Psychological Review*, 1977: 231-259.

³⁷ Kuhn, D. "Children and adults as intuitive scientists." *Psychological Review*, 1989: 674-689.

³⁸ Kunda, Z. "The case for motivating reasoning." *Psychological Bulletin*, 1990: 480-498

³⁹ Perkin, D. N., M. Farady, and B. Bushey. "Everyday reasoning and the roots of intelligence." In *Informal reasoning and education*, by J.F. Voss, D.N. Perkins and J.W. Segal, 83-105. New Jersey : Erlbaum, 1991.

⁴⁰ Shavitt, S. "The role of attitude objects in attitude formation." *Journal of Experimental Social Psychology*, 1990: 124-148.

⁴¹ Edwards, K., and W. von Hippel. "Hearts and minds: The priority of affective versus cognitive factors in person perception." *Personality and Social Psychology Bulletin*, 1995: 996-101.

⁴² Asch, S. "Studies of independence and conformity: A minority of one against a unanimous majority." *Psychological Monographs*, 1956.

⁴³ Bargh, J.A., S. Chaiken, P. Raymond, and C. Hymes. "The automatic evaluation effect: Unconditionally automatic activation with a pronunciation task." *Journal of Experimental Social Psychology*, 1996: 185-210.

⁴⁴ Davis, J.L., and C.E. Rusbult. "Attitude alignment in close relationships." *Journal of Personal and Social Psychology*, 2001: 65-84.

⁴⁵ Newcomb, T.M. *Personality and social change: Attitude formation in a student community*. New York: Dryden, 1943.

⁴⁶ Sherif, M. "A study of some social factors in perception." *Archives of Psychology*, 1935: 187.

⁴⁷ Haidt, J., and F. Bjorklund. "Social Intuitionists answer six questions about moral psychology." In *Moral Psychology, Volume 2: The Cognitive Science of Morality: Intuition and Diversity.*, edited by W. Sinnott-Armstrong, 181-217. Cambridge: MIT Press, 2008.

-
- ⁴⁸ Wilson, T.D., S. Lindsey, and T. Schooler. "A model of dual attitudes." *Psychological Review*, 2000: 101-126.
- ⁴⁹ Selman, R. "The relation of role taking to the development of moral judgment in children." *Child Development*, 1971: 79-91
- ⁵⁰ Piaget, J. *The moral judgment of the child* (M. Gabain, Trans.). New York: Free Press, 1932/1965
- ⁵¹ Kohlberg, L. "The cognitive-development approach to socialization." In *Handbook of socialization theory and research*, by D. A. Goslin, 347-480. Chicago: Rand McNally, 1969.
- ⁵² Kohlberg, L. "From is to ought: How to commit the naturalistic fallacy and get away with it in the study of moral development." In *Cognitive development and epistemology*, by T Mschel, 151-235. New York: Academic Press, 1971.
- ⁵³ Haidt, Jonathan. "The Emotional Dog and it's Rational Tail: A Social Intuitionist Approach to Moral Judgments." *Psychological Review*, 2001: [819].
- ⁵⁴ Zajonc, R.B. "Feeling and thinking: Preferences need no inferences." *American Psychologist*, 1980: 151-175.
- ⁵⁵ Bargh, J.A., S. Chaiken, P. Raymond, and C. Hymes. "The automatic evaluation effect: Unconditionally automatic activation with a pronunciation task." *Journal of Experimental Social Psychology*, 1996: 185-210.
- ⁵⁶ Fazio, R.H., D.M. Sanbonmatsu, M.C. Powell, and F.R. Kardes. "On the automatic evaluation of attitudes." *Journal of Personality and Social Psychology*, 1986: 229-238.
- ⁵⁷ Murphy, S.T., and R.B. Zajonc. "Affect, cognition, and awareness: Affective priming with optimal and suboptimal stimulus exposures." *Journal of Personality and Social Psychology*, 1993: 723-729.
- ⁵⁸ Hermans, D., J. De Houwer, and P. Eelen. "The affective priming effect: Automatic evaluation of evaluative information in memory." *Cognition and Emotion*, 1994: 515-533.
- ⁵⁹ Haidt, Jonathan. "The Emotional Dog and it's Rational Tail: A Social Intuitionist Approach to Moral Judgments." *Psychological Review*, 2001: 819
- ⁶⁰ Ibid
- ⁶¹ Albright, L., D. A. Kenny, and T.E. Malloy. "Consensus in personality judgments at zero acquaintance." *Journal of Personality and Social Psychology*, 1988: 387-395.
- ⁶² Devine, P.G. "Stereotypes and prejudice: Their automatic and controlled components." *Journal of Personality and Social Psychology*, 1989: 5-18.
- ⁶³ Chaiken, S. "The heuristic model of persuasion." In *Social influence: The Ontario symposium*, by M.P. Zanna, J.M. Olson and C.P. Herman, 3-39. Hillsdale: Erlbaum, 1987.
- ⁶⁴ Davis, J.L., and C.E. Rusbult. "Attitude alignment in close relationships." *Journal of Personal and Social Psychology*, 2001: 65-84

-
- ⁶⁵ Lord, C.G., L. Ross, and M.R. Lepper. "Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence." *Journal of Personality and Social Psychology*, 1979: 2098-2109.
- ⁶⁶ Haidt, Jonathan. "The Emotional Dog and its Rational Tail: A Social Intuitionist Approach to Moral Judgments." *Psychological Review*, 2001: 821
- ⁶⁷ Chen, S., D. Schechter, and S Chaiken. "Getting at the truth or getting along: Accuracy-versus impression-motivated heuristic and systematic processing." *Journal of Personality*, 1996: 262-275.
- ⁶⁸ Festinger, L. *A theory of cognitive dissonance*. Stanford: Stanford University Press, 1957.
- ⁶⁹ Wicklund, R.A., and J. W. Brehm. *Perspectives on cognitive dissonance*. Hillsdale: Erlbaum, 1976.
- ⁷⁰ Chaiken, S., R. Giner-Sorolla, and S. Chen. "Beyond accuracy: Defense and Impression motives in heuristic and systematic information processing." In *The psychology of action: Linking cognition and motivation to behavior*, by P.M. Gollwitzer and J.A. Bargh, 553-578. New York: Guilford, 1996.
- ⁷¹ Ibid
- ⁷² Lerner, M.J. "Evaluation of performance as a function of performer's reward and attractiveness." *Journal of Personality and Social Psychology*, 1965: 355-360
- ⁷³ Lerner, M.J., and D.T. Miller. "Just world research and the attribution process: Looking back and ahead." *Psychological Bulletin*, 1978: 1030-1051.
- ⁷⁴ Tetlock, P.E., O.V. Kristel, B. Elson, M. Green, and J. Lerner. "The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals." *Social Psychology*, 2000: 853-870.
- ⁷⁵ Solomon, S., J. Greenberg, and T. Pyszczynski. "A terror management theory of social behavior." In *Advances in experimental social psychology*, edited by M. Zanna, 93-159. San Diego: Academic Press, 1991.
- ⁷⁶ Rosenblatt, A, J Greenberg, S Solomon, T Pyszczynski, and D Lyon. "Evidence for terror management theory: The effects of morality salience on reactions to those who violate or uphold cultural values." *Journal of Personality and Social Psychology*, 1989: 681-690.
- ⁷⁷ Haidt, Jonathan. "The Emotional Dog and its Rational Tail: A Social Intuitionist Approach to Moral Judgments." *Psychological Review*, 2001: 822
- ⁷⁸ Nisbett, R.E., and S. Schacter. "Cognitive manipulation of pain." *Journal of Experimental Social Psychology*, 1966: 227-236.
- ⁷⁹ Nisbett, R.E., & Wilson, T.D. "Telling more than we know: Verbal reports on mental processes." *Psychological Review*, 1977: 231-259.
- ⁸⁰ Ibid.
- ⁸¹ Kunda, Z. "The case for motivating reasoning." *Psychological Bulletin*, 1990: 480-498.

-
- ⁸² Pyszczynski, T., and J. Greenberg. "Toward an integration of cognitive and motivational perspectives on social inference: A biased hypothesis-testing model." *Advances in Experimental Social Psychology*, 1987: 297-340.
- ⁸³ Zimbardo, P.G., S. Leberge, and L.D. Butler. "Psychophysiological consequences of unexplained arousal: A posthypnotic suggestion paradigm." *Journal of Abnormal Psychology*, 1993: 466-473.
- ⁸⁴ Kunst-Wilson, W.R., and R.B. Zajonc. "Affective discrimination of stimuli that cannot be recognized." *Science*, 1980: 557-558.
- ⁸⁵ Gazzaniga, M.S., J.E. Bogen, and R.W. Sperry. "Some functional effects of sectioning the cerebral commissures in man. ." *Proceedings of the National Academy of Sciences*, 1962: 1765-1769.
- ⁸⁶ Gazzaniga. *The social brain*. New York: Basic Books, 1985.
- ⁸⁷ Kohlberg, L. "From is to ought: How to commit the naturalistic fallacy and get away with it in the study of moral development." In *Cognitive development and epistemology*, by T Mschel, 151-235. New York: Academic Press, 1971.
- ⁸⁸ Haidt, Jonathan. "The Emotional Dog and it's Rational Tail: A Social Intuitionist Approach to Moral Judgments." *Psychological Review*, 2001: 814-834.
- ⁸⁹ Ibid, 823
- ⁹⁰ Blasi, A. "Bridging moral cognition and moral action: A critical review of the literature." *Psychological Bulletin*, 1980: 1-45. Quoted in: Haidt, Jonathan. "The Emotional Dog and it's Rational Tail: A Social Intuitionist Approach to Moral Judgments." *Psychological Review*, 2001: 823
- ⁹¹ Lapsley, D.K. *Moral Psychology*. Boulder: Westview, 1996.
- ⁹² Metcalf, J., and W. Mischel. "A hot/cool-system analysis of delay of gratification: Dynamics of willpower." *Psychological Review*, 1999: 3-19.
- ⁹³ Haidt, Jonathan. "The Emotional Dog and it's Rational Tail: A Social Intuitionist Approach to Moral Judgments." *Psychological Review*, 2001: 823
- ⁹⁴ Shoda, Y., W. Mischel, and P.K. Peake. "Predicting adolescent cognitive and self-regulatory competencies from preschool delay of gratification: Identifying diagnostic conditions." *Developmental Psychology*, 1990: 978-986.
- ⁹⁵ Kochanska, G., K. Murray, T.Y. Jaques, A.L. Koenig, and K.A. Vandegeest. "Inhibitory control in young children and its role in emerging internalization." *Child Development*, 1996: 490-507.
- ⁹⁶ Hirschi, T., and M.J. Hindelang. "Intelligence and delinquency: a revisionist view." *Sociological Review*, 1977: 571-87.
- ⁹⁷ Hart, D., and S. Fegley. "Prosocial behavior and caring in adolescence: Relations to self-understanding and social judgment." *Child Development*, 1995: 1346-1359.

-
- ⁹⁸ Colby, A. Damon, W. Some do care: Contemporary lives of moral commitment. New York: Free Press, 1992.
- ⁹⁹ Thoma, S.J., D. Narvaez, J. Rest, and P. Derryberry. "Does moral judgment development reduce to political attitudes or verbal ability? Evidence using the defining issues test." *Educational Psychology Review*, 1999: 325-341.
- ¹⁰⁰ Haidt, Jonathan. "The Emotional Dog and its Rational Tail: A Social Intuitionist Approach to Moral Judgments." *Psychological Review*, 2001: 824
- ¹⁰¹ Raine, A. "Antisocial behavior and psychophysiology: A biosocial perspective and prefrontal dysfunction hypothesis." In *Handbook of antisocial behavior*, edited by D.M. Stoff, J. Breiling and J. D. Maser, 289-304. New York: John Wiley, 1997.
- ¹⁰² Damasio, A. *Descartes's error: Emotion, reason, and the human brain*. New York: G.P. Putnam's Sons, 1994.
- ¹⁰³ Damasio, A. R., D. Tranel, and H. Damasio. "Individuals with sociopathic behavior caused by frontal damage fail to respond autonomically to social stimuli." *Behavioral Brain Research*, 1990: 81-94.
- ¹⁰⁴ Hare, R.D., and M.J. Quinn. "Psychopathy and autonomic conditioning." *Journal of Abnormal Psychology*, 1971: 223-235.
- ¹⁰⁵ Batson, C.D. "Prosocial motivation: Is it ever truly altruistic? ." *Advances in Experimental Psychology*, 1987: 65-122.
- ¹⁰⁶ Hoffman, M.L. "Development of prosocial motivation: Empathy and guilt." In *The Development of prosocial behavior*, edited by N. Eisenberg. New York: Academic Press, 1982.
- ¹⁰⁷ Cialdini, R.B. "Altruism or egoism? That is (still) the question." *Psychological Inquiry*, 1991: 124-126.
- ¹⁰⁸ Saltzstein, Herbert D., and Tziporah. Kasachkoff. "Haidt's Social Intuitionist Theory: A Psychological and Philosophical Critique." *Review of General Psychology*, 2004: 273-282
- ¹⁰⁹ Saltzstein, Herbert D., and Tziporah. Kasachkoff. "Haidt's Social Intuitionist Theory: A Psychological and Philosophical Critique." *Review of General Psychology*, 2004: 273-282.
- ¹¹⁰ Ibid 278
- ¹¹¹ Ibid 280-281
- ¹¹² Haidt, Jonathan. "The Emotional Dog Gets Mistaken for a Possum." *Review of General Psychology*, 2004: 283-290.
- ¹¹³ Edwards, K., and W. von Hippel. "Hearts and minds: The priority of affective versus cognitive factors in person perception." *Personality and Social Psychology Bulletin*, 1995: 996-101.
- ¹¹⁴ Shavitt, S. "The role of attitude objects in attitude formation." *Journal of Experimental Social Psychology*, 1990: 124-148.

-
- ¹¹⁵ Haidt, J., and F. Bjorklund. "Social Intuitionists answer six questions about moral psychology." In *Moral Psychology, Volume 2: The Cognitive Science of Morality: Intuition and Diversity.*, edited by W. Sinnott-Armstrong, 181-217. Cambridge: MIT Press, 2008.
- ¹¹⁶ Haidt, Jonathan. "The Emotional Dog and its Rational Tail: A Social Intuitionist Approach to Moral Judgments." *Psychological Review*, 2001: 827-828
- ¹¹⁷ Haidt dog possum 286
- ¹¹⁸ Haidt, Jonathan. "The Emotional Dog Gets Mistaken for a Possum." *Review of General Psychology*, 2004: 286
- ¹¹⁹ Greene, Joshua D., Nystrom, Leigh E. Sommerville, John M. Darley, and Jonathan D. Cohen. "An fMRI Investigation of Emotional Engagement in Moral Judgment." *Science*, 2001: 2105-2108.
- ¹²⁰ Greene, Joshua. "Dual-Process Morality and the Personal/Impersonal distinction: A reply to McGuire, Lagdon, Coltheart, and Mackenzie." *Journal of Experimental Social Psychology*, 2009: 581
- ¹²¹ Greene, J.D., R.B. Sommerville, L.E. Nystrom, J.M. Darley, and J.D. Cohen. "An fMRI investigation of emotional engagement in moral judgment." *Science*, 2001: 2105-2108.
- ¹²² Greene, Joshua D., Leigh E. Nystrom, Andrew D. Engell, John M. Darlet, and Jonathan D. Cohen. "The Neural Bases of Cognitive Conflict and Control in Moral Judgment." *Neuron*, 2004: 389-400.
- ¹²³ Botvinick, M.M., T.S. Braver, D.M. Barch, C.S. Carter, and J.D. Cohen. "Conflict monitoring and cognitive control." *Psychological Review*, 2001: 624-652.
- ¹²⁴ Greene, Joshua D., Sylvia A. Morelli, Kelly Lowenberg, and Leigh E. Cohen, Jonathan D. Nystrom. "Cognitive load selectively interferes with utilitarian moral judgment." *Cognition*, 2008: 1144-1154.
- ¹²⁵ Koenigs, M., L. Young, R. Adolphs, D. Tranel, F. Cushman, and M. Hauser. "Damage to the prefrontal cortex increases utilitarian moral judgments." *Nature*, 2007: 908-911.
- ¹²⁶ McGuire, Jonathan, Robyn Langdon, Max Coltheart, and Catriona Mackenzie. "A reanalysis of the personal/impersonal distinction in moral psychology research." *Journal of Experimental Social Psychology*, 2009: 577-580.
- ¹²⁷ Ibid, 578
- ¹²⁸ Ibid, 579
- ¹²⁹ Ibid, 580
- ¹³⁰ Greene, Joshua. "Dual-Process Morality and the Personal/Impersonal distinction: A reply to McGuire, Lagdon, Coltheart, and Mackenzie." *Journal of Experimental Social Psychology*, 2009.
- ¹³¹ Moll, Jorge, Roland Zahn, Ricardo Oliveira-Souza, Frank Krueger, and Jordan. Grafman. "The neural basis of human moral cognition." *Nature Reviews Neuroscience*, 2005: 799-809.
- ¹³² Ibid 805

-
- ¹³³ Ibid 805-806
- ¹³⁴ Ibid
- ¹³⁵ Moll, J., P.J. Eslinger, and R. Oliveria-Souza. "Frontopolar and anterior temporal cortex activation in a moral judgment task: preliminary functional MRI results in normal subjects." *Arq. Neuropsiquiatr*, 2001: 657-664
- ¹³⁶ Moll, J., Oliveria-Souzam R., I.E. Bramati, and J. Grafman. "Functional networks in emotional moral and nonmoral social judgments." *Neuroimage*, 2002: 696-703.
- ¹³⁷ Heekeren, H.R., I. Watenburger, H. Schmidt, H.P. Schwintowski, and A Villringer. "An fMRI study of simple ethical decision-making." *Neuroreport*, 2003: 1215-1219.
- ¹³⁸ Greene, J.D., R.B. Sommerville, L.E. Nystrom, J.M. Darley, and J.D. Cohen. "An fMRI investigation of emotional engagement in moral judgment." *Science*, 2001: 2105-2108.
- ¹³⁹ Greene, Joshua D., Leigh E. Nystrom, Andrew D. Engell, John M. Darlet, and Jonathan D. Cohen. "The Neural Bases of Cognitive Conflict and Control in Moral Judgment." *Neuron*, 2004: 389-400.
- ¹⁴⁰ Moll, Jorge, et al.. "The neural correlates of moral sensitivity: a functional magnetic resonance imaging investigation of basic moral emotions." *The Journal of Neuroscience*, 2002: 2730-2736.
- ¹⁴¹ Shin, Lisa M. et al.. "Activation of anterior paralimbic structures during guilt-related script driven imagery." *Society of Biological Psychiatry*, 2000: 43-50.
- ¹⁴² Takahashi, Hidehiko, Noriaki Yahata, Michihiko Koeda, Tetsuya Matsuda, Kunihiko Asai, and Yoshiro Okubo. "Brain activation associated with evaluative processes of guilt and embarrassment: an fMRI study." *New York Academy of Sciences*, 2008: 161-180.
- ¹⁴³ Moll, J., Oliveria-Souzam R., I.E. Bramati, and J. Grafman. "Functional networks in emotional moral and nonmoral social judgments." *Neuroimage*, 2002: 696-703.
- ¹⁴⁴ Moll, J., Oliveria-Souzam R., I.E. Bramati, and J. Grafman. "Functional networks in emotional moral and nonmoral social judgments." *Neuroimage*, 2002: 696-703.
- ¹⁴⁵ Shin, Lisa M. et al.. "Activation of anterior paralimbic structures during guilt-related script driven imagery." *Society of Biological Psychiatry*, 2000: 43-50.
- ¹⁴⁶ Casbeer, William D. *Processes and Moral Emotions*. Vols. Volume 3: The Neuroscience of Morality: Emotion, Brain Disorders, and Development, in *Moral Psychology*, 19-24. Cambridge: MIT Press, 2008.
- ¹⁴⁷ Hynes, Catherine A. *Morality, Inhibition, and Propositional Content*. Vols. Volume 3: The Neuroscience of Morality: Emotion, Brain Disorders, and Development, in *Moral Psychology*, 25-30. Cambridge: MIT Press, 2008.
- ¹⁴⁸ Moll, Jorge, Mirella L.M.F. Paivva, Roland Zahn, and Jordan Grafman. *Moral Psychology*. Vols. Volume 3: The Neuroscience of Morality: Emotion, Brain Disorders, and Development., in *Moral Psychology*, edited by Waltern Sinnott-Armstrong, 31-34. Cambridge: MIT Press, 2008.

¹⁴⁹ Ibid

¹⁵⁰ Casbeer, William D. *Processes and Moral Emotions*. Vols. Volume 3: The Neuroscience of Morality: Emotion, Brain Disorders, and Development, in *Moral Psychology*, 19-24. Cambridge: MIT Press, 2008.

¹⁵¹ Ibid

¹⁵² Moll, Jorge, Mirella L.M.F. Paivía, Roland Zahn, and Jordan Grafman. *Moral Psychology*. Vols. Volume 3: The Neuroscience of Morality: Emotion, Brain Disorders, and Development., in *Moral Psychology*, edited by Waltern Sinnott-Armstrong, 31-34. Cambridge: MIT Press, 2008.

¹⁵³ Term coined by Gordon Davis, personal correspondence

¹⁵⁴ Parfit, Derek, and John Broome. "Reasons and Motivation." *Proceedings of the Aristotelian Society, Supplementary Volumes*, 1997: 99-146.

¹⁵⁵ Parfit, Derek. *On What Matters*. New York: Oxford University Press, 2011: 31-33

¹⁵⁶ Ibid, 50

¹⁵⁷ Pizzaro, David A., and Paul Bloom. "The Intelligence of the Moral Intuitions: Comment on Haidt (2001)." *Psychological Review*, 2003: 193-196.

¹⁵⁸ Lazarus, R.S. *Emotion and adaptation*. New York: Oxford University Press, 1991.

¹⁵⁹ Dandoy, A.C., and A.G. Goldstien. "The use of cognitive appraisal to reduce stress reactions: A replication." *Journal of Social Behavior and Personality*, 1990: 1275-1285.

¹⁶⁰ Betancourt, H. "An attribution-empathy model of helping behavior: Behavioral intentions and judgments of help-giving." *Personality and Social Psychology Bulletin*, 1990: 573-591.

¹⁶¹ Batson, C.D., J.L. Dyck, J.R. Brandt, J.G., Powell, A.L. Batson, M.R. McMaster, and C Griffit. "Five Studies testing new egoistic alternatives to the empathy-altruism hypothesis." *Journal of Personality and Social Psychology*, 1988: 52-57.

¹⁶² Pizzaro, David A., and Paul Bloom. "The Intelligence of the Moral Intuitions: Comment on Haidt (2001)." *Psychological Review*, 2003:194

¹⁶³ Dasgupta, N., and A. Greenwald. "The malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals." *Journal of Personality and Social Psychology*, 2001: 800-814.

¹⁶⁴ Rudman, L.A., R.D. Ashmore, and M.L. Gary. "'Unlearning' automatic biases: The malleability of implicit stereotypes and prejudice." *Journal of Personality and Social Psychology*, 2001: 856-868.

¹⁶⁵ Haidt, Jonathan. "The Emotional Dog Does Learn New Tricks: A Reply to Pizarro and Bloom ." *Psychological Review*, 2003: 197-198.

¹⁶⁶ Kuhn, D. "Children and adults as intuitive scientists." *Psychological Review*, 1989: 674-689.

-
- ¹⁶⁷ Perkins, D. N., M. Farady, and B. Bushey. "Everyday reasoning and the roots of intelligence." In *Informal reasoning and education*, by J.F. Voss, D.N. Perkins and J.W. Segal, 83-105. New Jersey : Erlbaum, 1991.
- ¹⁶⁸ Haidt, Jonathan. "The Emotional Dog Does Learn New Tricks: A Reply to Pizarro and Bloom ." *Psychological Review*, 2003: 198
- ¹⁶⁹ Moll, Jorge, and Ricardo de. Oliveira-Souza. "Moral judgments, emotions and the utilitarian brain." *Trends in Cognitive Sciences*, 2007: 319-321.
- ¹⁷⁰ Koenigs, M., L. Young, R. Adolphs, D. Tranel, F. Cushman, and M. Hauser. "Damage to the prefrontal cortex increases utilitarian moral judgments." *Nature*, 2007: 908-911.
- ¹⁷¹ Damasio A. *Looking for Spinoza: Joy, Sorrow and the Feeling Brain*. Orlando: Harcourt, 2003.
- ¹⁷² Koenigs, M., L. Young, R. Adolphs, D. Tranel, F. Cushman, and M. Hauser. "Damage to the prefrontal cortex increases utilitarian moral judgments." *Nature*, 2007: 908-911.
- ¹⁷³ Greene, Joshua D. "Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains." *Trends in Cognitive Science*, 2007: 322-323.
- ¹⁷⁴ Ibid. 322

Bibliography

-
- Albright, L., D. A. Kenny, and T.E. Malloy. "Consensus in personality judgments at zero acquaintance." *Journal of Personality and Social Psychology*, 1988: 387-395.
- Asch, S. "Studies of independence and conformity: A minority of one against a unanimous majority." *Psychological Monographs*, 1956.
- Ayer, A.J. *Language, Truth and Logic*. Oxford: Oxford University Press, 1936.
- Bargh, J. "The four horsemen of automaticity: Awareness, efficiency, intention, and control in social cognition." In *Handbook of social cognition*, by J. R. S. Wyer and T. K. Srull, 1-40. Hillsdale: Erlbaum, 1994.
- Bargh, J.A., S. Chaiken, P. Raymond, and C. Hymes. "The automatic evaluation effect: Unconditionally automatic activation with a pronunciation task." *Journal of Experimental Social Psychology*, 1996: 185-210.
- Batson, C.D. "Prosocial motivation: Is it ever truly altruistic? ." *Advances in Experimental Psychology*, 1987: 65-122.
- Batson, C.D., J.L. Dyck, J.R. Brandt, J.G., Powell, A.L. Batson, M.R. McMaster, and C Griffit. "Five Studies testing new egoistic alternatives to the empathy-altruism hypothesis." *Journal of Personality and Social Psychology*, 1988: 52-57.
- Berger, P.L., & Luckman, T. *The social construction of reality*. New York: Doubleday, 1967.
- Betancourt, H. "An attribution-empathy model of helping behavior: Behavioral intentions and judgments of help-giving." *Personality and Social Psychology Bulletin*, 1990: 573-591.
- Blackburn, Simon. "Critical Notice of Allan Gibbard's Wise Choices, Apt Feelings." *Ethics*, 1992.
- Blasi, A. "Bridging moral cognition and moral action: A critical review of the literature." *Psychological Bulletin*, 1980: 1-45.
- Botvinick, M.M., T.S. Braver, D.M. Barch, C.S. Carter, and J.D. Cohen. "Conflict monitoring and cognitive control." *Psychological Review*, 2001: 624-652.
- Casbeer, William D. *Processes and Moral Emotions*. Vols. Volume 3: The Neuroscience of Morality: Emotion, Brain Disorders, and Development, in *Moral Psychology*, 19-24. Cambridge: MIT Press, 2008.
- Chaiken, S. "The heuristic model of persuasion." In *Social influence: The Ontario symposium*, by M.P. Zanna, J.M. Olson and C.P. Herman, 3-39. Hillsdale: Erlbaum, 1987.
- Chaiken, S., R. Giner-Sorolla, and S. Chen. "Beyond accuracy: Defense and Impression motives in heuristic and systematic information processing." In *The psychology of action: Linking cognition and motivation to behavior*, by P.M. Gollwitzer and J.A. Bargh, 553-578. New York: Guilford, 1996.

-
- Chen, S., D. Schechter, and S. Chaiken. "Getting at the truth or getting along: Accuracy-versus impression-motivated heuristic and systematic processing." *Journal of Personality*, 1996: 262-275.
- Cialdini, R.B. "Altruism or egoism? That is (still) the question." *Psychological Inquiry*, 1991: 124-126.
- Colby, A. Damon, W. *Some do care: Contemporary lives of moral commitment*. New York: Free Press, 1992.
- Damasio, A. *Descartes's error: Emotion, reason, and the human brain*. New York: G.P. Putnam's Sons, 1994.
- . *Looking for Spinoza: Joy, Sorrow and the Feeling Brain*. Orlando: Harcourt, 2003.
- Damasio, A. R., D. Tranel, and H. Damasio. "Individuals with sociopathic behavior caused by frontal damage fail to respond autonomically to social stimuli." *Behavioral Brain Research*, 1990: 81-94.
- Dandoy, A.C., and A.G. Goldstien. "The use of cognitive appraisal to reduce stress reactions: A replication." *Journal of Social Behavior and Personality*, 1990: 1275-1285.
- Dasgupta, N., and A. Greenwald. "The malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals." *Journal of Personality and Social Psychology*, 2001: 800-814.
- Davis, J.L., and C.E. Rusbult. "Attitude alignment in close relationships." *Journal of Personality and Social Psychology*, 2001: 65-84.
- Davis, J.L., and C.E. Rusbult. "Attitude alignment in close relationships." *Journal of Personal and Social Psychology*, 2001: 65-84.
- Devine, P.G. "Stereotypes and prejudice: Their automatic and controlled components." *Journal of Personality and Social Psychology*, 1989: 5-18.
- Edwards, K., and W. von Hippel. "Hearts and minds: The priority of affective versus cognitive factors in person perception." *Personality and Social Psychology Bulletin*, 1995: 996-101.
- Ekman, Paul. "An Argument for Basic Emotions." *Cognition and Emotion*, 1992: 169-200.
- Fazio, R.H., D.M. Sanbonmatsu, M.C. Powell, and F.R. Kardes. "On the automatic evaluation of attitudes." *Journal of Personality and Social Psychology*, 1986: 229-238.
- Festinger, L. *A theory of cognitive dissonance*. Stanford: Stanford University Press, 1957.
- Gazzaniga. *The social brain*. New York: Basic Books, 1985.

-
- Gazzaniga, M.S., J.E. Bogen, and R.W. Sperry. "Some functional effects of sectioning the cerebral commissures in man. ." *Proceedings of the National Academy of Sciences*, 1962: 1765-1769.
- Gibbard, Allan. *Reconciling Our Aims*. New York: Oxford University Press, 2008.
- . *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. New York: Oxford University Press, 1990.
- Greene, J.D., R.B. Sommerville, L.E. Nystrom, J.M. Darley, and J.D. Cohen. "An fMRI investigation of emotional engagement in moral judgment." *Science*, 2001: 2105-2108.
- Greene, Joshua D. "Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains." *Trends in Cognitive Science*, 2007: 322-323.
- Greene, Joshua D., Leigh E. Nystrom, Andrew D. Engell, John M. Darlet, and Jonathan D. Cohen. "The Neural Bases of Cognitive Conflict and Control in Moral Judgment." *Neuron*, 2004: 389-400.
- Greene, Joshua D., Nystrom, Leigh E. Sommerville, John M. Darley, and Jonathan D. Cohen. "An fMRI Investigation of Emotional Engagement in Moral Judgment." *Science*, 2001: 2105-2108.
- Greene, Joshua D., Sylvia A. Morelli, Kelly Lowenberg, and Leigh E. Cohen, Jonathan D. Nystrom. "Cognitive load selectively interferes with utilitarian moral judgment." *Cognition*, 2008: 1144-1154.
- Greene, Joshua. "Dual-Process Morality and the Personal/Impersonal distinction: A reply to McGuire, Lagdon, Coltheart, and Mackenzie." *Journal of Experimental Social Psychology*, 2009.
- Haidt, J., and F. Bjorklund. "Social Intuitionists answer six questions about moral psychology." In *Moral Psychology, Volume 2: The Cognitive Science of Morality: Intuition and Diversity*, edited by W. Sinnott-Armstrong, 181-217. Cambridge: MIT Press, 2008.
- Haidt, Jonathan. "The Emotional Dog and its Rational Tail: A Social Intuitionist Approach to Moral Judgments." *Psychological Review*, 2001: 814-834.
- Haidt, Jonathan. "The Emotional Dog Does Learn New Tricks: A Reply to Pizarro and Bloom. ." *Psychological Review*, 2003: 197-198.
- Haidt, Jonathan. "The Emotional Dog Gets Mistaken for a Possum." *Review of General Psychology*, 2004: 283-290.
- Hare, R.D., and M.J. Quinn. "Psychopathy and autonomic conditioning." *Journal of Abnormal Psychology*, 1971: 223-235.

-
- Hart, D., and S. Fegley. "Prosocial behavior and caring in adolescence: Relations to self-understanding and social judgment." *Child Development*, 1995: 1346-1359.
- Heekeren, H.R., I. Watenburger, H. Schmidt, H.P. Schwintowski, and A Villringer. "An fMRI study of simple ethical decision-making." *Neuroreport*, 2003: 1215-1219.
- Hermans, D., J. De Houwer, and P. Eelen. "The affective priming effect: Automatic evaluation of evaluative information in memory." *Cognition and Emotion*, 1994: 515-533.
- Hirschi, T., and M.J. Hindelang. "Intelligence and delinquency: a revisionist view." *Sociological Review*, 1977: 571-87.
- Hoffman, M.L. "Development of prosocial motivation: Empathy and guilt." In *The Development of prosocial behavior*, edited by N. Eisenberg. New York: Academic Press, 1982.
- Hynes, Catherine A. *Morality, Inhibition, and Propositional Content*. Vols. Volume 3: The Neuroscience of Morality: Emotion, Brain Disorders, and Development, in *Moral Psychology*, 25-30. Cambridge: MIT Press, 2008.
- Johnson, Gregory. "Theories of Emotion." *Internet Encyclopedia of Philosophy*. June 9, 2009. <http://www.iep.utm.edu/emotion/> (accessed July 20, 2012).
- Joyce, Richard. *The Myth of Morality*. New York: Cambridge University Press, 2001.
- Kauppinen, Antti. "The Rise and Fall of Experimental Philosophy." *Philosophical Explorations*, 2007: 95-118.
- Kohlberg, L. "From is to ought: How to commit the naturalistic fallacy and get away with it in the study of moral development." In *Cognitive development and epistemology*, by T Mschel, 151-235. New York: Academic Press, 1971.
- Knobe, Joshua, and Shaun Nichols. "An Experimental Philosophy Manifesto." In *Experimental Philosophy*, by Shaun Nichols, 3-17. New York: Oxford University Press, 2008.
- Kochanska, G., K. Murray, T.Y. Jaques, A.L. Koenig, and K.A. Vandegeest. "Inhibitory control in young children and its role in emerging internalization." *Child Development*, 1996: 490-507.
- Koenigs, M., L. Young, R. Adolphs, D. Tranel, F. Cushman, and M. Hauser. "Damage to the prefrontal cortex increases utilitarian moral judgments." *Nature*, 2007: 908-911.
- Kohlberg, L. "The cognitive-development approach to socialization." In *Handbook of socialization theory and research*, by D. A. Goslin, 347-480. Chicago: Rand McNally, 1969.
- Kuhn, D. "Children and adults as intuitive scientists." *Psychological Review*, 1989: 674-689.

-
- Kunda, Z. "The case for motivating reasoning." *Psychological Bulletin*, 1990: 480-498.
- Kunst-Wilson, W.R., and R.B. Zajonc. "Affective discrimination of stimuli that cannot be recognized." *Science*, 1980: 557-558.
- Lapsley, D.K. *Moral Psychology*. Boulder: Westview, 1996.
- Lazarus, R.S. *Emotion and adaptation*. New York: Oxford University Press, 1991.
- Lerner, M.J. "Evaluation of performance as a function of performer's reward and attractiveness." *Journal of Personality and Social Psychology*, 1965: 355-360.
- Lerner, M.J., and D.T. Miller. "Just world research and the attribution process: Looking back and ahead." *Psychological Bulletin*, 1978: 1030-1051.
- Lord, C.G., L. Ross, and M.R. Lepper. "Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence." *Journal of Personality and Social Psychology*, 1979: 2098-2109.
- McGrath, Matthew. "Propositions." *Stanford Encyclopedia of Philosophy*. December 19, 2005. <http://plato.stanford.edu/entries/propositions/> (accessed July 20, 2012).
- McGuire, Jonathan, Robyn Langdon, Max Coltheart, and Catriona Mackenzie. "A reanalysis of the personal/impersonal distinction in moral psychology research." *Journal of Experimental Social Psychology*, 2009: 577-580.
- Metcalf, J., and W. Mischel. "A hot/cool-system analysis of delay of gratification: Dynamics of willpower." *Psychological Review*, 1999: 3-19.
- Moll, J., Oliveria-Souzam R., I.E. Bramati, and J. Grafman. "Functional networks in emotional moral and nonmoral social judgments." *Neuroimage*, 2002: 696-703.
- Moll, J., P.J. Eslinger, and R. Oliveria-Souza. "Frontopolar and anterior temporal cortex activation in a moral judgment task: preliminary functional MRI results in normal subjects." *Arq. Neuropsiquiatr*, 2001: 657-664.
- Moll, Jorge, and Ricardo de. Oliveira-Souza. "Moral judgments, emotions and the utilitarian brain." *Trends in Cognitive Sciences*, 2007: 319-321.
- Moll, Jorge, et al.. "The neural correlates of moral sensitivity: a functional magnetic resonance imaging investigation of basic moral emotions." *The Journal of Neuroscience*, 2002: 2730-2736.
- Moll, Jorge, Mirella L.M.F. Paivia, Roland Zahn, and Jordan Grafman. *Moral Psychology*. Vols. Volume 3: The Neuroscience of Morality: Emotion, Brain Disorders, and Development., in *Moral Psychology*, edited by Waltern Sinnott-Armstrong, 31-34. Cambridge: MIT Press, 2008.

-
- Moll, Jorge, Roland Zahn, Ricardo Oliveira-Souza, Frank Krueger, and Jordan. Grafman. "The neural basis of human moral cognition." *Nature Reviews Neuroscience*, 2005: 799-809.
- Murphy, S.T., and R.B. Zajonc. "Affect, cognition, and awareness: Affective priming with optimal and suboptimal stimulus exposures." *Journal of Personality and Social Psychology*, 1993: 723-729.
- Newcomb, T.M. *Personality and social change: Attitude formation in a student community*. New York: Dryden, 1943.
- Nisbett, R.E., & Wilson, T.D. "Telling more than we know: Verbal reports on mental processes." *Psychological Review*, 1977: 231-259.
- Nisbett, R.E., and S. Schacter. "Cognitive manipulation of pain." *Journal of Experimental Social Psychology*, 1966: 227-236.
- Parfit, Derek. *On What Matters*. New York: Oxford University Press, 2011.
- Parfit, Derek, and John Broome. "Reasons and Motivation." *Proceedings of the Aristotelian Society, Supplementary Volumes*, 1997: 99-146.
- Perkins, D. N., M. Farady, and B. Bushey. "Everyday reasoning and the roots of intelligence." In *Informal reasoning and education*, by J.F. Voss, D.N. Perkins and J.W. Segal, 83-105. New Jersey : Erlbaum, 1991.
- Piaget, J. *The moral judgment of the child (M. Gabain, Trans.)*. New York: Free Press, 1932/1965.
- Pizzaro, David A., and Paul Bloom. "The Intelligence of the Moral Intuitions: Comment on Haidt (2001)." *Psychological Review*, 2003: 193-196.
- Prinz, Jesse. *The Emotional Construction of Morals*. New York: Oxford University Press, 2008.
- Pyszczynski, T., and J. Greenberg. "Toward an integration of cognitive and motivational perspectives on social inference: A biased hypothesis-testing model." *Advances in Experimental Social Psychology*, 1987: 297-340.
- Raine, A. "Antisocial behavior and psychophysiology: A biosocial perspective and prefrontal dysfunction hypothesis." In *Handbook of antisocial behavior*, edited by D.M. Stoff, J. Breiling and J. D. Maser, 289-304. New York: John Wiley, 1997.
- Rosenblatt, A, J Greenberg, S Solomon, T Pyszczynski, and D Lyon. "Evidence for terror management theory: The effects of morality salience on reactions to those who violate or uphold cultural values." *Journal of Personality and Social Psychology*, 1989: 681-690.
- Rudman, L.A., R.D. Ashmore, and M.L. Gary. "'Unlearning' automatic biases: The malleability of implicit stereotypes and prejudice." *Journal of Personality and Social Psychology*, 2001: 856-868.

-
- Saltzstein, Herbert D., and Tziporah. Kasachkoff. "Haidt's Social Intuitionist Theory: A Psychological and Philosophical Critique." *Review of General Psychology*, 2004: 273-282.
- Selman, R. "The relation of role taking to the development of moral judgment in children." *Child Development*, 1971: 79-91.
- Shavitt, S. "The role of attitude objects in attitude formation." *Journal of Experimental Social Psychology*, 1990: 124-148.
- Sherif, M. "A study of some social factors in perception." *Archives of Psychology*, 1935: 187.
- Shin, Lisa M., and et al.. "Activation of anterior paralimbic structures during guilt-related script driven imagery." *Society of Biological Psychiatry*, 2000: 43-50.
- Shoda, Y., W. Mischel, and P.K. Peake. "Predicting adolescent cognitive and self-regulatory competencies from preschool delay of gratification: Identifying diagnostic conditions." *Developmental Psychology*, 1990: 978-986.
- Smith, Michael. *The Moral Problem*. New York: Blackwell Publishing, 1994.
- Solomon, S., J. Greenberg, and T. Pyszczynski. "A terror management theory of social behavior." In *Advances in experimental social psychology*, edited by M. Zanna, 93-159. San Diego: Academic Press, 1991.
- Stevenson, C.L. *Ethics and Language*. New Haven: Harvard University Press, 1944.
- Takahashi, Hidehiko, Noriaki Yahata, Michihiko Koeda, Tetsuya Matsuda, Kunihiro Asai, and Yoshiro Okubo. "Brain activation associated with evaluative processes of guilt and embarrassment: an fMRI study." *New York Academy of Sciences*, 2008: 161-180.
- Tetlock, P.E., O.V. Kristel, B. Elson, M. Green, and J. Lerner. "The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals." *Social Psychology*, 2000: 853-870.
- Thoma, S.J., D. Narvaez, J. Rest, and P. Derryberry. "Does moral judgment development reduce to political attitudes or verbal ability? Evidence using the defining issues test." *Educational Psychology Review*, 1999: 325-341.
- Van Roojen, Mark. "Moral Cognitivism vs. Non-Cognitivism." *Stanford Encyclopedia of Philosophy*. 2009. <http://plato.stanford.edu/entries/moral-cognitivism/> (accessed July 20, 2012).
- Wicklund, R.A., and J. W. Brehm. *Perspectives on cognitive dissonance*. Hillsdale: Erlbaum, 1976.

Wilson, T.D., S. Lindsey, and T. Schooler. "A model of dual attitudes." *Psychological Review*, 2000: 101-126.

Zajonc, R.B. "Feeling and thinking: Preferences need no inferences." *American Psychologist*, 1980: 151-175.

Zimbaro, P.G., S. Laberge, and L.D. Butler. "Psychophysiological consequences of unexplained arousal: A posthypnotic suggestion paradigm." *Journal of Abnormal Psychology*, 1993: 466-473.