



uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES
ET POSTDOCTORALES



FACULTY OF GRADUATE AND
POSTDOCTORAL STUDIES

Jason L. Millar

AUTEUR DE LA THÈSE / AUTHOR OF THESIS

M.A. (Philosophy)

GRADE / DEGRÉ

Department of Philosophy

FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Affecting Moral Judgment

TITRE DE LA THÈSE / TITLE OF THESIS

Ian Kerr

DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

Patrice Philie

Andrew Sneddon

David Castle

Gary W. Slater

Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

Affecting Moral Judgment

Jason L. Millar

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
In partial fulfillment of the requirements
For the MA degree in Philosophy

Department of Philosophy
Faculty of Arts
University of Ottawa

© Jason L. Millar, Ottawa, Canada, 2008



Library and
Archives Canada

Bibliothèque et
Archives Canada

Published Heritage
Branch

Direction du
Patrimoine de l'édition

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file *Votre référence*
ISBN: 978-0-494-48486-9
Our file *Notre référence*
ISBN: 978-0-494-48486-9

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.

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

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

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


Canada

Table of Contents

Acknowledgments	iv
1. A Model of Moral Rationalism	1
Introduction	1
Moral Rationalism	6
Plato's Tripartite Divisions of the Soul.....	6
Mill and Pleasure Derived From Higher Human Faculties	13
Kant's Impartial Rational Spectator	17
The Model	21
The Model of Moral Rationalism.....	22
Questioning the Model on Psychological Grounds.....	26
Contemporary Echoes of the Model	29
Singer's Objection to the Relevance of Moral Psychology	30
Legal Reasoning and Adjudication.....	31
Rational Principles in Biomedical Contexts	36
Conclusion	37
2. The Psychological Delineations of Moral Rationalism	39
Introduction	39
Investigating the Theoretical Separation of Emotions and Moral Reasoning	41
The <i>Is-Ought</i> Barrier to Empirical Research.....	42
The Empirical Turn: Moral Psychology and Neurology.....	49
Implications For the Model of Moral Rationalism	61
Conclusion	67
3. Towards A Model of Moral Psychological Multidimensionality	72

Introduction	72
Two Questions For Moral Psychology	73
Rethinking the Model of Moral Rationalism	75
Core Moral Judgment and Sentimental Rules	76
Sentimental Rules and the Model of Moral Rationalism	84
A Model of Moral Psychological Multidimensionality.....	91
Moral Emotional Engagement is Gradated.....	93
Moral Emotional Engagement is Multidimensional	97
Conclusions.....	100
4. Postscript - Looking to Future Research	102
5. Bibliography	106

Acknowledgments

I owe a great deal of thanks to my thesis supervisor, Ian Kerr, who at the time of writing holds the Canada Research Chair in Ethics, Law & Technology at the University of Ottawa. He has worked tirelessly to create an academic environment within which his students are limited only by their own efforts. On top of that, his mentorship has been invaluable. In addition, Andrew Sneddon deserves thanks for introducing me to so many of the topics in philosophy that have captured my imagination and managed to hold it for a good while. Thanks also to David Castle, who holds the Canada Research Chair in Science and Society at the University of Ottawa. All three individuals provided insightful comments on this thesis and improved it enormously.

I owe a special thanks to Michelle Morley, whose patience and support has made this entire journey possible.

1. A Model of Moral Rationalism

Introduction

Moral philosophers often describe hypothetical situations in which a moral agent is required to choose among two or more difficult courses of action. These moral dilemmas generally highlight salient features of the particular situation to demonstrate how the agent can appeal to a set of normative principles in judging between courses of action. Normative principles are meant to be applied as a means of providing a moral justification for their decision. Choosing among options is often difficult because it involves selecting among competing normative principles, that is, those proffered by certain moral philosophers who disagree over which set of principles is the better one upon which the agent ought to base his ultimate decision. In many moral dilemmas the agent experiences what might be described as an emotional reaction, for example anger or disgust, which is at odds with those principles, further complicating an already difficult deliberation. Though the agent might feel drawn to consider a course of action in direct response to the emotional reaction rather than one based on normative, or rational, principles, many philosophers from Plato onward have discounted that as a legitimate moral response. The general argument against the relevance of emotions in moral reasoning is that they are functionally distinct from reason, particularly

untrustworthy in moral deliberations, and ought therefore to be ignored if one is to reach a proper rational moral judgment.¹

Historical figures in moral philosophy, such as Plato, Immanuel Kant and John Stuart Mill, argue against the legitimacy of appeasing the emotions in moral judgment.² For them, good reasons are the legitimate path to good moral judgments. Indeed, those philosophers do not merely argue for the primacy of reasons as justificatory components in moral deliberation, they argue for the near exclusivity of reason as a justificatory component in moral deliberation. That is, they propose that moral agency ought to involve justificatory appeals to reasons devoid of emotional interference. As Moore put it, "the main object of Ethics, as a systematic science, is to give correct *reasons* for thinking that this or that is good."³ Our emotional mind, if it presents itself, ought to be subservient to the rational justifications offered in favour of a particular moral decision, where those rational justifications ought not to find their basis in appeals to our emotional response. By denying a role for the emotions in moral deliberation and judgment, those moral philosophers suggest that moral deliberation

¹ I will discuss this generalization in more detail throughout this chapter.

² Plato, *Phaedrus*, trans. Christopher Rowe (New York, New York: Penguin Books, 2005).; Plato, *Republic*, trans. G.M.A. Grube and C. D.C. Reeve (Indianapolis, Indiana: Hackett Publishing Company, 1992).; Immanuel Kant, "The Foundations of the Metaphysics of Morals," in *Ethical Theory: A Concise Anthology*, ed. Heimer Geirsson and Margaret R. Holmgren, 113-130 (Peterborough, Ontario: Broadview Press, 2000).; John Stuart Mill, "Utilitarianism," in *Ethical Theory: A Concise Anthology*, ed. Heimer Geirsson and Margaret R. Holmgren, 86-94 (Peterborough, Ontario: Broadview Press, 2000).

³ G.E. Moore, *Principia Ethica* (Cambridge: Cambridge University Press, 1965), 6.

can occur along emotional-free lines of reasoning.⁴ Hence, those philosophers offer rational arguments meant to provide reasons for doing such and such in a given situation, regardless of whether the reasons and resulting normative principles come in conflict with an agent's emotions.

Throughout this thesis I will refer to any theory in moral philosophy that includes the premise that moral deliberation and judgments ought to be premised on reasons to the near exclusion of emotions as a *moral rationalist* theory, or more generally as the category of theories called *moral rationalism*.⁵

In this chapter I will draw from historical theories in moral philosophy, specifically those developed by Plato, Mill and Kant, to explicate the roots of moral rationalism. I will also describe a common element shared by theirs, and other subsequent moral rationalist theories. Plato, Mill and Kant are three of the most influential of the historical figures in moral philosophy. A common element shared by their theories could have interesting implications for contemporary theoretical and applied ethics owing to that historical influence. I will argue that the common element those historical moral rationalist theories share is their reliance on a particular model

⁴ This argumentative move seems to originate with Kant. I will discuss it in more detail later in this chapter.

⁵ Richard Taylor, *Good and Evil: A New Direction* (London: Collier-Macmillan Ltd., 1970). From what I can tell Taylor coined the term Moral Rationalism, describing it as the particular model that sees the will guided by reason rather than the emotions. He attributes this model to Plato, Aristotle and Kant. Moral rationalism stands in contrast to other philosophical theories that propose a legitimate role for the emotions to play in anchoring moral judgment. Those metaethical theories, such as Emotivism, will not be discussed in this thesis. My focus is on the particular treatment given to the emotions in moral rationalist theories.

of moral psychology that is implicitly described in the theories. By discussing examples of contemporary moral rationalist applications, I will argue that the historical model does indeed resonate in current applied contexts. I will show that other philosophers have recently challenged moral rationalist theories on psychological grounds. Since moral rationalists have proffered a particular model of moral psychology I will argue that their model is open to challenges on empirical psychological grounds.

First I should clarify some of the terms I will be using throughout this thesis. I will often refer to a process of moral deliberation and the judgment that a moral agent reaches as the result of a moral deliberation. The existing literature reveals that there is an ongoing debate about what it means to *judge* and what a *judgment* is with respect to a *deliberation*.⁶ For example, Haidt defines moral deliberation as a process of “conscious mental activity that consists of transforming given information about people in order to reach a moral judgment.”⁷ Moral judgments are defined as “evaluations (good vs. bad) of the actions or character of a person that are made with respect to a set of virtues held to be obligatory by a culture or subculture.”⁸ By pointing out the definitions I mean to draw attention to the idea that moral deliberation and moral judgment can be theoretically distinguished, the latter following from the former in informed moral agency. I do not intend to engage that debate and, as such, will remain agnostic with regards to particular philosophical or psychological conceptions of them. I do not

⁶ See J Haidt, "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment," *Psychological Review* 108, no. 4 (2001): 814-834 for a good example of an analysis of the relationship between moral deliberation and judgment.

⁷ Ibid, 818.

⁸ Ibid, 817.

believe that this will affect my argument, as I am merely interested in judgment as either a component, or end product, of a deliberative process. A distinction between the two should not impact the argument, nor would a lack of any distinction at all. In general, when I refer to deliberation or judgment, I will be referring to both the rational process of considering premises and arguments as a moral agent, and to the final decision that a moral agent makes after carefully considering various options, for example those options that are available in a hypothetical moral dilemma. For the sake of brevity I will tend to refer to both terms under the single term “judgment” or “moral judgment”.

I should also note that I am interested in addressing the kind of deliberation and judgment resulting from a deliberate process of thought. Therefore, I am not referring to the (often contradictory) moral intuitions of individuals caught in a series of real-life moral quandaries when I refer to deliberation or judgment. Though studies into our moral intuitions provide a rich literature on the kinds of decisions we make in our day-to-day activities, our intuitions seem to fall short in providing justifications during the kinds of moral deliberations demanded of us in the most difficult situations.

In addition, my use of relative terms in describing judgments as better, correct, proper, etc. should not be interpreted as epistemologically significant. This is not an argument about the epistemic truth of moral judgments. My indications that one judgment could be better, or more correct, than another is not meant to imply that it is. The language of moral philosophy necessitates those comparative terms, and in describing a philosopher’s project it is, I think, fair to describe them as each trying to

show why the other philosopher has got it wrong, or is less correct. The fact of the matter is as yet unknown, so any reference to relations between philosophers' judgments should not be interpreted as a statement about the fact of the matter. Where I mean to imply otherwise I will make it clear.

Moral Rationalism

There is a long tradition in moral philosophy of excluding the emotions from legitimate moral judgment. Plato, Mill and Kant developed moral rationalist theories that share this common element of exclusion. But in order for a philosopher to successfully argue for such exclusion he must first argue that there is a distinction to be made between whatever is illegitimate in moral judgment and what counts as legitimate. In the case of moral judgment, the general distinction is drawn based on what moral rationalists refer to as "rational capacities" and "non-rational capacities" or the emotions. Plato was one of the earliest proponents of this distinction. His basic distinction between rational aspects of the mind and non-rational (emotional) aspects of the mind set the stage for various influential moral rationalists including John Stuart Mill, who argued for a general rational principle upon which to justify moral judgments, and Immanuel Kant, who argued that a rational adherence to duty alone constitutes good judgment.

Plato's Tripartite Divisions of the Soul

Plato's theory of moral judgment derive from his detailed descriptions of the human soul, set out in the dialogues *Phaedrus* and *Republic*. Each of these descriptions provides an account of the soul as divided into distinct parts, each performing unique functions, as well as a description of how the parts are different in kind from each other. Plato's

approach to analysing moral agency is a good starting point for understanding moral rationalism in general because he sets out some basic distinctions that carry through in those theories proposed by Mill and Kant, and resonate in contemporary applied contexts.⁹

In the *Phaedrus* Plato says the immaterial soul is most god-like in its composition because it contains the intellect as one of its parts. Thus, it alone is capable of acquiring true knowledge; from true knowledge the soul derives happiness.¹⁰ His treatment of the soul in the *Republic* is similar, though it focuses on the divisions of the soul as they correspond to those contained in a just city.¹¹ Examining the nature of the soul is therefore Plato's method of determining how it is that humans are capable of living the good life.

The model Plato chooses as a representation of the soul in the *Phaedrus* "resemble[s] the combined power of a winged team of horses and their charioteer".¹² Though there are three players, Plato is clear that there are only two *kinds* of thing that compose the soul:

⁹ These applied contexts will be discussed in turn later in this chapter.

¹⁰ Plato, *Phaedrus*, (247c-d) (n.2).

¹¹ Plato, *Republic*, (Book IV) (n.2).

¹² Plato, *Phaedrus*, (246a8-9) (n.2).

“In each of us there are two kinds of thing which rule and lead us, which we follow wherever they may lead, the one an inborn desire for pleasures, the other an acquired judgment that aims at the best.”¹³

Our charioteer and the first horse are similar to the gods in that they are “noble and good, and of similar stock”, while the second horse “is of the opposite stock, and opposite in its nature”.¹⁴ The charioteer is the “intellect...nourished and made happy by gazing on what is true”.¹⁵ Its partner, the first horse, is said to be good because it is “a lover of honour when joined with restraint and a sense of shame, and a companion of true glory, needing no whip, responding to spoken orders alone”.¹⁶ The second horse is not good, a “companion of excess and boastfulness...deaf, and hardly yielding to whip and goad together.”¹⁷

Given the inherent competing interests possessed by its different parts, Plato likens the soul to a battlefield.¹⁸ The way an individual lives, his moral stock, depends on which of the combatants gains control,

“Now when judgment leads us by reason towards the best and is in control, its control over us has the name of restraint; when desire drags us irrationally

¹³ Ibid, (237d-238a).

¹⁴ Ibid, (246a9-b4).

¹⁵ Ibid, (247d).

¹⁶ Ibid, (253d6-e1). Rowe explains that ‘spoken orders’ refers to reason in this passage.

¹⁷ Ibid, (253e3-5).

¹⁸ Ibid, (253-4).

towards pleasures and has established rule within us, its rule is called by the name of excess.”¹⁹

To lead the good life one must acquire that which is best—true knowledge. Acquiring true knowledge begins with the soul ‘seeing’ “being which really is, which is without colour or shape, intangible, observable by the steersman of the soul alone.”²⁰ This kind of ‘being’ is a reference to Plato’s ‘Forms’ or ‘Ideas’, perfect models from which the stuff of our everyday lives is imperfectly copied.²¹ Included among the Forms are “justice itself...self-control, [and] knowledge.”²² For the soul to catch a glimpse of what really is, the charioteer and the good horse must be in control, guided by reason, such that they can rise up above earthly things toward which the bad horse draws us. The more rational we are, the better we live,

“The chariots of the gods travel easily, being well balanced and easily controlled, while the rest do so with difficulty; for the horse that is partly bad weighs them down, inclining them towards the earth with its weight, if any of the charioteers has not trained him well.”²³

According to Plato’s theory, living a better life, that is, being more rational, involves relegating the influence of the bad horse to that of the intellect and good horse combined. Consider a passage in which he describes the soul’s reaction to a beautiful

¹⁹ Ibid, (237d-238a).

²⁰ Ibid, (247c7-8). Rowe points out that ‘seeing’ is used metaphorically.

²¹ Ibid, pp.79-80. See author’s note 101.

²² Ibid, (247d6-8).

²³ Ibid, (247b2-5).

person. The bad horse lurches violently, pulling the pair towards the “delights of sex”.²⁴ Suddenly the charioteer remembers the true “nature of beauty” and “sees it standing together with self-control on a holy pedestal” causing him to feel shame and throw himself back on the reins to stop any forward progress.²⁵ This pattern repeats in an epic internal battle between reason and desire until passion is left bloodied and, humbled by reason, gives in to its guidance.²⁶

Plato’s treatment of the soul in the *Republic* expresses the same general divisions set out in the *Phaedrus*. The soul is described as containing three parts: a rational part, an appetitive part and a spirited part.²⁷ As was the case with the charioteer, the rational part is said to have wisdom in that it possesses “good judgment” derived from a particular kind of knowledge.²⁸ In this case the rational part of the soul is said to be able to attain knowledge by studying philosophy.²⁹ Of the other two parts the appetitive corresponds to the uncooperative horse. It is the part of the soul that “lusts, hungers, thirsts, and gets excited by other appetites” and is “the irrational appetitive part, companion of certain indulgences and pleasures”.³⁰ As for the third part of the soul that seems to correspond to the cooperative horse, Plato appears to clarify its function and relationship to the rational part. He says that the spirited part is that by

²⁴ Ibid, (254a9).

²⁵ Ibid, (254b6-7).

²⁶ Ibid, (254e).

²⁷ Plato, *Republic*, (440e-441a) (n.2).

²⁸ Ibid, (428-429a).

²⁹ Ibid, (502c-509c).

³⁰ Ibid, (439d).

which we get “angry” or “disgusted.”³¹ Despite the apparent similarity to the appetites he explicitly distinguishes the spirited part from the appetitive part:

“We often notice in other cases that when appetite forces someone contrary to rational calculation, he reproaches himself and gets angry with that in him that’s doing the forcing, so that of the two factions that are fighting a civil war, so to speak, spirit allies itself with reason.”³²

In doing so Plato asserts that there is a difference in kind between anger and disgust on the one hand, and the appetitive desires on the other. Yet he maintains that for a soul to be just, the rational part must play the part of ruler:

“[It is] appropriate for the rational part to rule, since it is really wise and exercises foresight on behalf of the whole soul, and for the spirited part to obey it and be its ally.”³³

Thus we have two descriptions of the soul, each offered up as a means of describing how it is that a person can lead a good life. A human soul contains three parts. Two of those are similar in kind in that each displays non-rational animalistic responses, while the third is rational. One non-rational part is base, drawn towards desires of the flesh, while the other is nobler, acting almost as a conscience through its reactions of anger and disgust. Each part is different in function from the others. The non-rational parts are ruled—one willingly, the other not—by the rational part, which

³¹ Ibid, (439e).

³² Ibid, (440a-b).

³³ Ibid, (441e).

is noblest, most god-like and is alone capable of delivering good judgments through the acquisition of true knowledge.

A general feature emerges from Plato's model of the soul, as it relates to leading a good life, which is of particular importance to the development of an understanding of moral rationalism's treatment of the emotions. A stark delineation is drawn between reason and what I will refer to as "the emotions" (or simply, "emotions"), which include anger, disgust, and the desires of the flesh like hunger, thirst, lust, and so on.³⁴ This can be seen both in the physical representations of the parts of the soul contained in Plato's two accounts, and in the qualitative description of each of those parts.

According to Plato's descriptions each part of the soul either has a discrete body, as in the case of the horses and charioteer, or is considered as a separate part as in the *Republic*. Each of reason and the emotions also has its own set of unique traits, none of which seem to overlap in the descriptions. Their functional roles are also used as delineators. Plato insists that reason alone is capable of guiding the soul towards leading the good life, while emotions act either to frustrate the intellect in its attempt to steer the soul towards the good, as is the case with desires of the flesh, or merely act as an irrational sidekick to the good charioteer, as is the case with anger and disgust. Both of the irrational parts of the soul require steering if we are to avoid living lives of corruption and injustice.³⁵ Reason and the emotions are also described as acting in

³⁴ For the time being it is enough to distinguish between a thoroughly rational part of the soul and other parts that display no rational functionality.

³⁵ Ibid, (Book VIII). Plato describes how the just city is prone to corruption and vice. In doing so he suggests that the human soul is similarly constituted.

direct opposition to one another. Reason is trustworthy as it naturally leads the soul towards the good and that which nourishes it, namely true knowledge, while desires of the flesh are corrupt, leading the soul towards what is bad, causing it to “waste away and perish”.³⁶ Without the rational charioteer at the reins our irrational spirited part would fight a constant battle with our appetitive part, unable to effectively train the appetites and incapable of knowing in which direction the good life lies.

Plato’s model of moral judgment suggests that a moral agent ought to distinguish between his emotional from his rational mental influences in moral judgment. In addition, he suggests that it is indeed possible to do so. This assertion sets the stage for a general treatment of the emotions by moral rationalists. It is restated to varying degrees in the theories proposed by Mill and Kant, and echoes of it can be found in contemporary philosophical theories and in applied contexts.

Mill and Pleasure Derived From Higher Human Faculties

“The creed which accepts as the foundation of morals ‘utility’ or the ‘greatest happiness principle’ holds that actions are right in proportion as they tend to promote happiness; wrong as they tend to produce the opposite of happiness. By happiness is intended pleasure and the absence of pain; by unhappiness, pain and the privation of pleasure.”³⁷

A complete account of Mill’s *Utilitarianism* might indicate that his ethical theory is not as thoroughly rationalistic as Plato’s. Indeed, it contains elements that are much

³⁶ Plato, *Phaedrus*, (246e5) (n.2).

³⁷ John Stuart Mill, "Utilitarianism," 86 (n.2).

more sympathetic to the emotions.³⁸ However, his philosophy also contains descriptive and prescriptive elements reminiscent of more rationalistic theorists. Given the importance of Mill's theory, especially in applied contexts, it is useful to examine it in order to ascertain whether or not the similarities are problematic.

Mill's Greatest Happiness Principle, on first reading, seems thoroughly hedonistic in its conception, placing the promotion of pleasure at the justificatory root of all moral judgment. But he explains that interpreting his principle merely as promoting hedonism would be a mistake. Interpreting it correctly requires understanding the special role that reason plays in generating pleasures unique to humans.³⁹

Mill argues that there are different kinds of pleasures, in that the quality of some outweighs the quantity of others in the opinion of humans. "Of two pleasures," he says, "if there be one to which all or almost all who have experience of both give a decided preference, irrespective of any feeling of moral obligation to prefer it, that is the more desirable pleasure."⁴⁰ Of all the pleasures, he argues "a manner of existence which employs the higher faculties" so far outweighs an existence "promising...the fullest allowance of a beast's pleasures" in quality, as to "render [the latter], in comparison, of small account."⁴¹ In arguing that the higher faculties produce a greater happiness than beastly pleasures, Mill declares, "the pleasures derived from the higher faculties to be

³⁸ Andrew Sneddon, "Feeling Utilitarian," *Utilitas* 15, no. 3 (November 2003): 330-352.

³⁹ John Stuart Mill, "Utilitarianism," 86 (n.2).

⁴⁰ *Ibid*, 87.

⁴¹ *Ibid*, 87-88

preferable *in kind*", that is they are not merely more intense or greater in quantity than the lower pleasures.⁴² And so he claims "it is better to be a human being dissatisfied than a pig satisfied; better to be Socrates dissatisfied than a fool satisfied."⁴³

Mill secures the Greatest Happiness Principle by arguing that exercising the higher faculties produces greater happiness in the individual, then by arguing that happiness is contagious,

"If it may possibly be doubted whether a nobler character is always the happier for its nobleness, there can be no doubt that it makes other people happier, and that the world in general is immensely a gainer by it."⁴⁴

From this he suggests that the utilitarian standard of happiness is the collective happiness of "the whole sentient creation."⁴⁵

But what, in particular, are the higher faculties? This last quote hierarchically organizes existences from greatest to least, starting with Socrates's, moving to the lesser of the fool's until we reach the lowest existence of the pig. A ranking of this sort suggests, at the very least, that the higher faculties include the capacity for reason. Furthermore, if we buy into the notion of a rank ordering of faculties, the greater the capacity for reason, the better the life lived by the individual.

⁴² Ibid, 89.

⁴³ Ibid, 88.

⁴⁴ Ibid, 90.

⁴⁵ Ibid.

Like Plato, Mill hints that the higher faculties function independently from the lower ones. He does it first by suggesting that one can slip into an existence devoid of the higher faculties. Mill describes the “capacity for the nobler feelings” as a “tender plant” in need of constant upkeep through intellectual pursuits.⁴⁶ Left unattended the higher faculties are “easily killed”, leaving the individual to “sink into indolence and selfishness...not because they deliberately prefer [the lower faculties], but because they are either the only ones to which they have access or the only ones that they are any longer capable of enjoying.”⁴⁷

The second suggestion pointing to the independent functioning of higher and lower faculties can be seen in his Greatest Happiness Principle, the application of which suggests that good moral judgments are better reached by diminishing the influence of the lower faculties. It prescribes something like a calculation to determine which action an agent ought to perform, which in turn is based solely on the amount of happiness that act will generate. An agent’s particular emotional character or responses does not feature explicitly in either his description of the utilitarian ideal or of happiness, leaving us to wonder what particular role they play according to Mill. However, given the rational formulation of the Principle it seems that for an agent to refer to his emotions for guidance during moral deliberation would be for the agent to satisfy one of the lower, beastly faculties. Those lower faculties constantly threaten to dull his intellect

⁴⁶ Ibid, 89.

⁴⁷ Ibid.

and pull him down into a life less worth living; the greatest happiness is gained through exercise of the higher faculties to as great an extent as is possible.

Kant's Impartial Rational Spectator

Of the philosophers who have contributed greatest to the tradition of moral rationalism Immanuel Kant is clearer than most in his rejection of the moral worth of emotions. His analysis of moral agency begins and ends with an analysis of what he terms the *good will*, the only thing in the world that "can be called good without qualification".⁴⁸

Determining whether an agent's will is good or not boils down to determining whether an agent acts out of a sense of duty or not, where duties are determined by means of the application of a thoroughly rational principle, often referred to as the *Categorical Imperative*.⁴⁹

A classic example of Kant's moral rationalist approach involves a moral dilemma in which an agent answers his door and is confronted by an axe murderer.⁵⁰ The murderer explains that he has come to kill the agent's family and asks if the family is currently home. If he discovers they are home, he will certainly kill them. If not, he will go away. Knowing that his family is home, the agent is faced with the dilemma whether to lie to the murderer and save his family, or tell the truth, securing their death. Being of good will he understands that he must will to do whatever duty requires of him; the

⁴⁸ Immanuel Kant, "The Foundations of the Metaphysics of Morals," 113 (n.2).

⁴⁹ See Chapter 14 of Frederick S.J. Copleston, *A History of Philosophy: Modern Philosophy, Part II, Kant*, Vol. VI (Garden City, N.Y.: Image Books, 1960) for a good and full explication of Kant's moral philosophy.

⁵⁰ Immanuel Kant, "On a Supposed Right to Lie because of Philanthropic Concerns," in *Grounding for the Metaphysics of Morals*, trans. James Wesley Ellington (Hackett Publishing, 1993), 65.

good will acts in accordance with duty for duty's sake regardless of selfish inclinations. In order to determine what his particular duty is with respect to lying to the axe murderer, the agent applies the Categorical Imperative, which states "Act as if the maxim of thy action were to become by thy will a universal law of nature".⁵¹ In this case the agent applies it by asking 'which of my two options could I will to become a universal natural law for all with respect to lying?' Given that it seems impractical to will that everyone lie as a universal law of nature, it becomes clear to the agent that he has a general duty to tell the truth. Being of good will he recognizes his duty to tell the truth and acts in accordance with it "*because duty requires*", despite his selfish inclinations to lie and save his family.⁵²

Kant's version of moral rationalism makes it impossible for a will to be good if its actions are based on the agent's emotional, or selfish, response. *Inclinations*, a term he uses to refer to non-rational impulses that often compete with reason to influence the will, can result from any number of emotional responses. This means that some inclinations will agree with an agent's duties while others will not. Acting to satisfy an inclination that is not derived from duty, or "done from duty"⁵³ as Kant puts it, has "no intrinsic unconditional value" and can at best presuppose a good will.⁵⁴ Kant's explanation for this stems from his metaphysics, in which objects of the will are material in nature whereas objects of reason, the natural "governor of our will", are

⁵¹ Ibid, 120.

⁵² Ibid, 115. (Kant's emphasis.)

⁵³ Ibid.

⁵⁴ Ibid, 113-114.

formal.⁵⁵ Thus, the formal objects of reason are the only things that can properly derive the will's moral worth. As Kant puts it,

“The will stands between its *a priori* principle, which is formal, and its *a posteriori* spring, which is material, as between two roads, and as it must be determined by something, it follows that it must be determined by the formal principle of volition when an action is done from duty, in which case the material principle has been withdrawn from it.”⁵⁶

Thus, the moral worth of an action done from duty is derived “from the maxim from which it is determined, [i.e.] the *principle of volition* by which the action has taken place, without regard for any object of desire.”⁵⁷ A result of this view is that the consequences of an action can never determine the moral worth of the action, nor can inclinations, the objects of which are material:

“Now an action done from duty must wholly exclude the influence of inclination, and with it every object of the will except objectively the *law*, and subjectively *pure respect* for this practical law, and consequently the maxim that I should follow this law even to the thwarting of all my inclinations.”⁵⁸

So if the agent is of good will, the axe murderer is delivered up his victims.

⁵⁵ Ibid, 114.

⁵⁶ Ibid, 116. (Kant's emphasis.)

⁵⁷ Ibid. (Kant's emphasis.)

⁵⁸ Ibid, 116-117. (Kant's emphasis.) The law referred to here is duty, which, on Kant's view, is naturalistic with respect to the good will.

Another formulation of the categorical imperative points to the social implications of Kant's moral rationalism.⁵⁹ Kant states it as follows, "act as to treat humanity, whether in your own person or in that of any other, always at the same time as an end, and never merely as a means."⁶⁰ This particular formulation is derived from Kant's argument that there is a supreme categorical imperative that applies specifically to the human will. Recognizing that "rational nature exists as an end in itself"⁶¹ rationally binds us to accept that any rational being is an end in itself, and must be treated accordingly.⁶² This formulation of the categorical imperative can be applied to the axe murderer case to determine the right course of action. Recognizing that to lie to the murderer would be to treat him as a means to one's selfish ends (the selfish end to live a life with one's family), one is duty bound to tell the truth. Construed another way, to lie would be to deprive the axe murderer of information required to make an informed and reasonable decision regarding his actions. Either way, the family is delivered up to the axe murderer.

Like Plato, and to a lesser extent Mill, Kant's analysis of moral foundations sees reason as the only natural governor of the will. A good will is guided by reason, which in turn derives actions from duty while ignoring (excluding) non-rational inclinations;

⁵⁹ Theodore C. Denise, Nicholas P. White and Sheldon P. Peterfreund, "Duty and Reason," in *Great Traditions in Ethics*, ed. Theodore C. Denise, Nicholas P. White and Sheldon P. Peterfreund, 143-156 (Belmont, CA: Wadsworth, 2002).

⁶⁰ Frederick S.J. Copleston, *A History of Philosophy: Modern Philosophy, Part II, Kant*, 120 (n.49).

⁶¹ Immanuel Kant, "The Foundations of the Metaphysics of Morals," 123-124 (n.2).

⁶² Frederick S.J. Copleston, *A History of Philosophy: Modern Philosophy, Part II, Kant*, 120 (n.49).

the good will functions as an “impartial rational spectator” duty bound to laws derived from the various formulations of the categorical imperative.⁶³ Inclinations encompass the emotions, among other desires, since the emotions are considered non-rational. Thus, the emotions are excluded from actively participating as justificatory elements of moral judgments. Kant’s complete exclusion of the emotions goes further even than Plato. Plato seems to accept that certain emotions, disgust or guilt for instance, are different than desires like lust or hunger, and are the natural followers of the will. Like Mill, Kant formulates a rational law meant to function as a touchstone for moral judgment; in deliberation an agent focuses her rational capacities on the formula to determine the correct action, which may or may not satisfy whatever non-rational elements are affecting her—the formulas are explicitly designed to exclude emotion from consideration.

Kant follows the lead of Plato and Mill in arguing that there is a distinction to be made between reason and the emotions, and similarly in arguing that an agent ought to distinguish between his rational and non-rational mental influences in moral judgment.

The Model

It is difficult to overstate the influence that Plato, Mill and Kant have had on the history of Western moral philosophy. Their ethical theories have been adapted for application in medical practise and research⁶⁴, academic research and public policymaking⁶⁵ to

⁶³ Immanuel Kant, "The Foundations of the Metaphysics of Morals," 113 (n.2).

⁶⁴ See Chapter 1 of Walter Glannon, *Biomedical Ethics* (New York: Oxford University Press, 2005). Glannon points out that utilitarianism and deontology are the two

name but a few wide-ranging applied contexts. It is of practical importance to ensure that any theories with such applied import are accurate so that the judgments they tend to prescribe have, at the very least, a good chance at standing the test of scrutiny, at best a chance of getting something right.

The Model of Moral Rationalism

Analysing this group of theories under a single model may seem problematic given the theoretical differences between say, Mill's consequentialist Utilitarianism on the one hand, and Kantian Deontology on the other, which explicitly invalidates the consideration of consequences in moral deliberation. But despite each of them arriving at a unique, perhaps mutually exclusive⁶⁶, theory of moral rationalism, each

primary moral theories used to justify judgments in biomedical ethics, both for medical practise and research.

⁶⁵ In Canada, the guidelines for determining the permissibility of scientific studies, particularly those involving the use of human or animal subjects, are set by the Canadian Council on Animal Care (CCAC). Universities comply with those standards even though they are not required to by statute. Researchers applying for federal funding must comply with the Tri-Council Policy Statement *Ethical Conduct for Research Involving Humans*, a public policy statement designed to ensure the import of theoretical ethical principles into practical decision-making. The CCAC guidelines can be found at their website, "List of CCAC Guidelines," *Canadian Council on Animal Care*, February 22, 2008, http://www.ccac.ca/en/CCAC_Programs/Guidelines_Policies/GDLINES/Guidelis.htm (accessed August 28, 2008). Also see Chapter 10 of Eike-Heiner W. Kluge, *Readings in Biomedical Ethics: A Canadian Focus*, 3rd Edition, ed. Eike-Heiner W. Kluge (Toronto: Pearson Prentice Hall, 2005) for a reproduction of, and discussion on, the Tri-Council Policy Statement. The Statement can also be found online, see National Council on Ethics in Human Research, "Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans," *National Council on Ethics in Human Research*, 1999, http://www.ncehr-cnerh.org/english/code_2/ (accessed June 16, 2008).

⁶⁶ Glannon, *Biomedical Ethics*, (n.63) argues that contemporary deontological theories have been friendlier to consequences than Kant, having been modified to treat consequences as secondary to duties. Likewise, he argues that consequentialism, having

philosopher treats reason and the emotions in a similar manner when prescribing their proper roles in moral judgment.

That similarity opens up an avenue for analysis. We can develop a general model of moral rationalism, insofar as we are discussing the respective roles of reason and the emotions, by focusing on a common element between the various theories and treating it as a primary feature of the model. We can then question that aspect of the theories, that is, we can question whether moral rationalists have portrayed the relative roles of reason and the emotions accurately, and determine whether the model is in need of revision, however slight, based on the results of the analysis. The slightest of revisions could have practical import.

Providing examples of the potential practical import of hypothetical revisions to a moral theory is a complex proposition. But looking to our legal system, with a particular focus on the nature of legal judgments, might highlight one context in which inaccuracies in the moral rationalist model could have practical implications. In our legal system both judges and jurors are asked to apply legal principles in rendering judgments on particular cases, the assumption being that they can function much like the impartial rational spectators explicitly proposed by Kant, implicitly by Plato and to some extent by Mill. As thoroughly rational agents judges and jurors are charged with hearing the details of a case in order to judge the arguments on the merits of the case alone. One aspect of this caricature is that the judge or juror ought not to allow their

been modified into rule-consequentialism, allows for a consideration of duties so long as consequences remain the primary concern.

emotions to interfere and cloud their judgments. But suppose the impartial rational spectator turned out to be an accident of philosophical oversight, and that we found instead that we had good reason to reassess the caricature in favour of a more nuanced individual whose emotions seemed to be more subtly entangled with his most rational judgments, this despite his best efforts to remain impartial. Could this affect our expectations or design of the legal system?

Any practical concerns arising as a result of an analysis of the moral rationalists' theories will remain purely speculative until we determine the theoretical lines along which to perform an analysis. That determination hinges in part on the question, "How should we interpret the moral rationalists' general treatment of reason and the emotions?"

We have seen that each of Plato, Mill and Kant describe the moral relevance of both reason and the emotions, arguing that living the good life consists in understanding that each ought to function in such and such a manner with respect to moral judgment. Reason is functionally defined with respect to the emotions; it is the moral ruler, the commander-in-chief that stands in charge of the will and governs our emotions, preventing them from tainting our judgment with shades of irrationality. Reason is described as if examining the emotions at arms length, determining whether they have any credible moral worth. Plato and Kant, and to some extent Mill, suggest that they do not—reason ought morally to ignore the non-rational emotions.

This is a common element that emerges from within the historical moral rationalist literature described in the preceding analysis, allowing us to better define

the model of moral rationalism. All of the above theories suggest that agents ought to distinguish between rational and non-rational (emotional) capacities in moral judgment. In other words, moral rationalists argue that despite the powerful influence the emotions often exert on our moral judgments, we ought morally to distinguish between our emotional and rational mental processes and defer to the rational aspects of our minds in moral judgment.

In making that argument moral rationalists are supposing a particular rudimentary model of moral psychology.⁶⁷ This is apparent first in the way they are providing a description of how the activities of the mind are starkly delineated and compartmentalized into distinct functional/qualitative parts, each of which can be implicated in moral judgment, and secondly in how the parts are hierarchically organized with respect to their roles in moral judgment.

A further aspect of the moral rationalist model must be considered, which is of particular practical importance. Kant is said to have claimed that any *ought* statement, a statement claiming that one *ought* to do so and so, implies that one *can* do it.⁶⁸ If we take that principle to apply equally to one's ability to reason in such and such a way, then moral rationalist claims about how one ought to judge uninfluenced by emotions imply that they *can* do so, that is, that an agent's rational capacities *can* be distinguished

⁶⁷ Richard Taylor, *Good and Evil: A New Direction*, (n.5).

⁶⁸ See Chapter 8 of Immanuel Kant, *Critique of Practical Reason*, ed. Mary Gregor, trans. Mary Gregor (Cambridge: Cambridge University Press, 1997), for the discussion surrounding the principle that ought implies can.

from his emotional ones in moral judgment.⁶⁹ Burton refers to the idea that the mind can be viewed as containing separate, distinguishable rational and non-rational actors as the “myth of the rational mind”.⁷⁰ Though he ultimately argues against it, he lauds its inherent appeal, “the repeated assertion of a rational mind sounds suspiciously like a disembodied mind capable of pure thought without inputs from bodily and mental sensations...the theory’s primary message—we can improve our reasoning by knowing when it has gone awry—is immensely appealing.”⁷¹

A rudimentary model of moral psychology can be derived from the historical moral rationalist treatment of reason and the emotions. It is the one I will examine throughout this thesis, and it can be sketched as containing a central psychological claim:

- i. Agents are capable of distinguishing between reason and emotion during moral judgment.

Questioning the Model on Psychological Grounds

I will refer to this moral rationalist model of moral psychology as the *moral rationalist model*, or simply as *the model*, from here forward.

⁶⁹ There is some controversy over whether or not this use of the principle that *ought implies can* is too strong, that Kant did not intend such a use. See Robert Stern, “Does ‘Ought’ Imply ‘Can’? And Did Kant Think It Does?,” *Utilitas* (Cambridge University Press) 16, no. 1 (March 2004): 42-61, for a discussion of this controversy. I will qualify its use by outlining what is implied by ‘can’ in the moral rationalist model.

⁷⁰ Robert A. Burton, *On Being Certain: Believeing You Are Right Even When You're Not* (New York: St. Martin's Press, 2008), 143.

⁷¹ *Ibid*, 145.

Philosophers have recognized the relevance of psychology to moral philosophical theories and have subsequently challenged moral theories along the lines of their psychological accuracy. In response to what she considers the “law conception” of ethics typified by both Mill and Kant, Anscombe claims,

“In present-day philosophy an explanation is required how an unjust man is a bad man, or an unjust action a bad one; to give such an explanation belongs to ethics; but it cannot even begin until we are equipped with a sound philosophy of psychology.”⁷²

Her assertion challenges philosophers to accept that moral judgments are in some way linked to psychological facts. She suggests those facts must be understood before accurate philosophical theories of moral judgment can be developed. More importantly those facts, once understood, would provide a touchstone for normative philosophical claims.

Along similar lines Flanagan argues, “all traditions of ethical thought are committed to a minimal sort of psychological realism.”⁷³ He further states the commitment in the form of a metaethical principle:

“PRINCIPLE OF MINIMAL PSYCHOLOGICAL REALISM [PMPR]: Make sure when constructing a moral theory or projecting a moral ideal that the character,

⁷² G.E.M. Anscombe, “Modern Moral Philosophy,” *Philosophy* 33, no. 124 (January 1958): 1-19, 4.

⁷³ Owen Flanagan, *Varieties of Moral Personality* (Cambridge, Mass.: Harvard University Press, 1991): 32.

decision processing, and behaviour prescribed are possible, or are perceived to be possible, for creatures like us.”⁷⁴

His PMPR sets clear limitations on the *ought* prescriptions that can be embedded in any moral theory, based on the possibilities of what *can* be achieved by creatures like us. Any theory found in violation of the PMPR would require revision, or suffer outright rejection in the event that its central claims were found psychologically untenable.

Taking seriously Anscombe’s challenge to provide an account of the philosophy of psychology as it relates to moral judgment, or Flanagan’s PMPR as a means of testing a moral theory, opens moral theories up to empirical investigation. An investigation of this sort could result in a more accurate description of our moral psychology based on the facts of the matter. If it turns out those facts about human psychology function as constraints on moral theories, for example if they point to limits to an agent’s ability to distinguish emotions from reason, then the facts matter. From this perspective it makes sense to attempt to model the empirical psychological claims contained in the predominant moral rationalist theories and test for their accuracy. By undertaking this kind of activity philosophers might identify gaps in the moral rationalist theories, correcting them as necessary. Such corrections would certainly be of historical and theoretical interest. However, there are echoes of the historical moral rationalist claims in contemporary moral rationalist theories, and those have been imported into applied contexts, adding urgency to the work. Before undertaking a fuller analysis of the model

⁷⁴ Ibid.

of moral rationalism, I will identify some of the areas in contemporary theoretical and applied moral philosophy that demonstrate this link to the historical trend.

Contemporary Echoes of the Model

I have mentioned that contemporary applied moral philosophy borrows heavily from historical moral rationalist theories.⁷⁵ A result of this is that contemporary theories and their historical predecessors might share the moral psychological supposition. If we find evidence of the model of moral rationalism, or of the principles derived from it, then we have reason to question the judgments prescribed by the contemporary theories, at least insofar as the judgments hinge on the accuracy of the model's central claim.

There are at least three contemporary applications that echo the model of moral rationalism. The first comes from a defense of moral rationalism argued by Peter Singer, in which he explicitly refers to the central claim of the model.⁷⁶ The second, which makes explicit reference to the central claim of the model, has already been discussed briefly and will be expanded here. It is the body of legal theory dealing with adjudication, particularly those theories that prescribe impartial applications of the law.⁷⁷ The third example can be found in the vast body of bioethical literature

⁷⁵ See notes 64-67.

⁷⁶ Peter Singer, "Ethics and Intuitions," *The Journal of Ethics* 9, no. 3/4 (October 2005): 331-352.

⁷⁷ *R. v. Krieger*, [2006] 2 S.C.R. 501, 2006 SCC 47.; John Finnis, "Natural Law and Legal Reasoning," in *Law and Morality: Readings in Legal Philosophy*, ed. David Dyzenhaus and Arthur Ripstein, 200-217 (Toronto: University of Toronto Press, 2001).; *R. v. Latimer*, [2001] 1 S.C.R. 3, 2001 SCC 1.

discussing applications of the categorical imperative or utilitarian principles to issues of patient care.⁷⁸ This last example relies on the model implicitly through its application of the Kantian and Utilitarian principles, which are based in turn on the central claim of the model of moral rationalism. An examination of the language used in arguing each of these groups of theories suggests that they all rely, to some extent, on the moral rationalist model for their efficacy.

Singer's Objection to the Relevance of Moral Psychology

Peter Singer, a noteworthy utilitarian, has responded directly to recent work published in moral psychology and moral philosophy that suggests a valid role for the emotions, or intuitions, in moral judgment. According to Singer, moral judgments can be separated into at least two categories: those that are based on intuitions arising as a result of "our evolutionary and cultural history", and those that are "more reasoned."⁷⁹ He argues that we can and ought to categorize judgments with this rubric to avoid validating the former kind, which would lead to moral skepticism.⁸⁰ Judgments of the former kind are not well-reasoned or well-principled judgments; being grounded on aspects of our culture or evolution they can be interpreted as inherently subject to change, unstable and anchored only by the (subjective) relative whims of particular individuals and groups. Judgments of the latter kind, on the other hand, "those that have a rational basis", promise to avoid the problem of moral skepticism because they

⁷⁸ Walter Glannon, *Biomedical Ethics* (n.64). Eike-Heiner W. Kluge, *Readings in Biomedical Ethics: A Canadian Focus*, 3rd Edition, ed. Eike-Heiner W. Kluge (n.65).

⁷⁹ Peter Singer, "Ethics and Intuitions," 350-351 (n.78).

⁸⁰ *Ibid*, 351.

are well reasoned—they adhere to the kinds of rational principles proffered by Kant and Mill.⁸¹ Those rational principles can be applied by any group in any culture to produce similar results, and thus are considered stable and relatively objective.

Singer's argument is a restatement of the central claim of the model of moral rationalism. He first identifies two kinds of judgments, one rational the other non-rational. Then he submits that we can and ought morally to undertake a process of categorization, which is itself a principled rational process, in order to separate the rational from non-rational judgments. The lead actor in the process of categorization is reason, like Plato's charioteer or Kant's impartial rational spectator, who stands back and judges the nature of judgments. Such a process relies on reason's ability to perform an emotion-free meta-judgment, which is much the same argument made by Plato, Mill and Kant. In order to avoid moral skepticism, individuals are asked to distinguish between rational and non-rational processes in evaluating moral judgments, the assumption being that they can, as a matter of fact, do so.

Legal Reasoning and Adjudication

Some legal theories of adjudication also explicitly assume the central claim of the model of moral rationalism in the way that judges and jurors are asked to apply the law, through impartial rational judgment, despite their feelings one way or another. I must sidestep a possible objection from the outset in using this example. In many (perhaps most, or all) cases, adjudication does not require judges or jurors to *morally* deliberate or render *moral* judgments; the deliberations undertaken and judgments

⁸¹ Ibid.

rendered in the law are considered *legal judgments*.⁸² From the legal perspective it seems to matter little whether or not agents are capable of deliberating or rendering judgments along impartial rational moral lines since they are not being asked to do so. But there are cases in which aspects of the legal judgment, demanded of judges or jurors by the law, closely resemble aspects of the kind of judgment demanded of moral agents by the model of moral rationalism. In those cases, it would be prudent to doubt the efficacy of the legal system to render the kinds of impartial judgments it prescribes, until such time as the accuracy of the central claim of the moral rationalist model can be settled one way or another.

Some aspects shared by moral and legal reasoning could be the determination of such thing as the degree of trustworthiness or blameworthiness, nature of character, or nature of the violation committed by any one of the many individuals involved in any aspect of a legal case. Each individual involved in a case, as well as the details of their actions and the information they provide, must be heard prior to being judged. But the kinds of deliberations and judgments involved in adjudicating on the various particulars of a case might not be reducible to legal ones devoid of moral import.

Finnis describes the event of listening to a lecturer to explicate what he refers to as the “four orders of reality with which human reason is concerned,” the last two of which create ambiguity between legal and moral reasoning:

⁸² John Finnis, "Natural Law and Legal Reasoning," (n.79).

“(3) One hears *the lecturer*, who (like the audience) is freely engaging in an activity and thereby participating in a human relationship: there is an order which one can bring into one’s own dispositions, choices, actions - one’s *praxis*, one’s doing, one’s *Existenz* – the “existential” order studied by some parts of psychology, by biography and the history of human affairs, and by moral and political philosophy. (4) One hears the *English language* and statements ordered by an expository or rhetorical technique, making and decoding the formalized symbols of a language and the less formalized but still conventional symbols and expressive routines of a cultural form and technique: the order one can intelligibly bring into matter which is subject to our power, so as to make objects such as phonemes, words, poems, boats, software, ballistic missiles and their inbuilt trajectories – the order of *poiesis*, of making of culture – studied in the arts and technologies, and in linguistics and rhetoric. (Corresponding to these four orders are four irreducibly distinct senses of “hearing.”)⁸³

The third order in his description corresponds roughly to the moral, social order, while the fourth corresponds roughly to a legal, or rule-based technical order of reason.

Problems arise in separating the aspects of moral reasoning from the aspects of legal reasoning since both are employed in the real-world process of adjudication. From this perspective, to describe legal deliberation as occurring only along impartial rational lines is to overlook the aspects of moral reasoning involved in the overall process of legal adjudication. Thus the moral aspects of deliberation are inherited by the overall

⁸³ Ibid, 205-206 (Finnis’s emphasis).

legal deliberation, along with its strengths and weaknesses of the central claim of the model of moral rationalism. Seen in this light, moral judgments do indeed factor as relevant aspects of legal adjudication.

The concept of *jury nullification* underscores the extent to which the idea of rendering impartial rational judgments is entrenched in the law, and further illustrates how the law could be affected by the validity of the moral rationalist model. Jury nullification occurs whenever a jury acquits a defendant “in disregard of the judge’s instructions and contrary to the jury’s findings of fact.”⁸⁴ Despite it sounding like a clear violation of important legal principles, such as following a judge’s instructions, or considering the weight of evidence in a case, jury nullification is an accepted practice in Canada, the United Kingdom and the United States. Though it is accepted, it is also discouraged and often controversial. The Supreme Court of Canada, in *R. v. Latimer*, stated that “guarding against jury nullification is a desirable and legitimate exercise for a trial judge; in fact a judge is required to take steps to ensure that the jury will apply the law properly.”⁸⁵ In *R. v. Morgentaler* an instance of jury nullification was appealed all the way to the Supreme Court of Canada, yet was upheld.⁸⁶ Most recently, the Supreme Court of Canada, in *R. v. Krieger*, upheld the principle of jury nullification in

⁸⁴ Merriam-Webster, 1982, <http://www.merriam-webster.com/dictionary/jury+nullification> (accessed July 29, 2008).

⁸⁵ *R. v. Latimer*, (n.79).

⁸⁶ *R. v. Morgentaler*, [1988] 1 S.C.R. 30.

cases where jurors could not, in good conscience, follow either the letter of the law or judges' instructions.⁸⁷

These cases exemplify a tension in the law between the concept of the impartial rational spectator, which is itself a restatement of the model of moral rationalism, and the potential for what might be considered a valid non-rational (possibly emotional) aspect of legal adjudication. On the one hand, the wording in *R. v. Latimer* makes it clear that jury nullification involves improper conduct on the part of the jury, which judges have a duty to guard against. One can imagine Kant as judge, making the argument to the jury that they ought morally to uphold the legal principles because only rational legal principles will produce judgments that can be applied universally. On the other hand, the specific language used in *R. v. Krieger* is an indication that the emotions might play a valid role in legal judgment, despite the requirement to apply rational legal principles prescribed by the law. The question of where one's conscience lies on any particular case is a moral question. One's conscientious objection could rest on a moral judgment based on rational moral principles. Or it could rest on a moral judgment derived otherwise, perhaps in part from some emotional component. The point is that the lack of specificity in *R. v. Krieger* suggests that the highest court in Canada is endorsing jury nullification that does not necessarily depend on an application of rational moral principles.

These considerations demonstrate that the accuracy of the model of moral rationalism could affect the practical aspects of the law in at least two ways. If the

⁸⁷ *R. v. Krieger*, (n.79).

model of moral rationalism is accurate, then one could argue that jury nullification is less permissible, especially in cases where jurors are unable to explain their objection on a rational basis. This would also bolster the practise of guarding against jury nullification and suggest that the wording in *R. v. Krieger* is problematic due to its moral ambiguity regarding the nature of valid conscientious objection. However, if the model of moral rationalism is found to be inaccurate, then *R. v. Krieger* gains weight. In addition, the practise of discouraging jury nullification and regarding it as improper would be more questionable. To what extent could judges reasonably demand (or expect) that jury members distinguish between their emotions and reason while deliberating on any particular case?

Rational Principles in Biomedical Contexts

Biomedical ethics is well populated by the deontological notion of irreducible duties like Kant's categorical imperative, and by examples of Mill's consequentialist Greatest Happiness Principle (albeit in slightly variable formulations).⁸⁸ Treating people as rational ends in themselves is the central justificatory tenet leading to the notion of the fully informed patient, which is a central principle upon which individuals are asked to base judgments regarding the design of medical research and patient-doctor relationships.⁸⁹ Similarly, a pluralistic take on Kant's basic idea of irreducible duties has resulted in principles like: the principle of autonomy and respect for persons, the principle of fidelity or best action, the principles of equality and justice, the

⁸⁸ Eike-Heiner W. Kluge, *Readings in Biomedical Ethics: A Canadian Focus*, 5-9 (n.65).

⁸⁹ Walter Glannon, *Biomedical Ethics*, (n.64).

principle of beneficence, and the principle of non-maleficence—each of which plays an important role in the judgments made daily in healthcare settings worldwide.⁹⁰

Similarly, consequentialist principles, very similar to Mill's, are often used to settle issues related to the allocation of scarce medical resources, and to aid in deliberations regarding triage.⁹¹

These various principles differ little from those originally proposed by Mill and Kant. As a result, they implicitly suppose the central claim of the model of moral rationalism. Difficult issues in healthcare are supposed to be settled in large through the rational application of these principles. This fact alone does not at all suggest that the principles suffer deep flaws. But it modestly suggests that there is reason to doubt the complete practical efficacy of the deliberations and judgments they prescribe, at least in proportion to any inaccuracies that might arise as a result of empirical investigation into the model of moral rationalism.

Conclusion

There is a long history in moral rationalist philosophy of arguing that an individual's rational capacities stand in a particular psychological relationship to her emotional ones. Reason and the emotions are described as functioning as if reason were a separate individual in the process of moral judgment, instructing or ignoring the various non-rational actors—desires, disgust, anger or guilt—who just happen to be present at the table. This description can be viewed as constituting a rudimentary

⁹⁰ Eike-Heiner W. Kluge, *Readings in Biomedical Ethics: A Canadian Focus*, 8-9 (n.65).

⁹¹ Walter Glannon, *Biomedical Ethics*, (n.64).

model of moral psychology, the central claim of which is that moral agents are capable of distinguishing between reason and emotion during moral judgment. A model of this kind is open to empirical investigation, the results of which may lead us to doubt the efficacy of the practical theories based on it. Contemporary echoes of the model can be heard in theories of applied moral philosophy, legal theory and theories in biomedical ethics. Inaccuracies in the model carry implications for those practical applications that rely on moral rationalist theories as a touchstone for moral judgment.

2. The Psychological Delineations of Moral Rationalism

Introduction

It should not be surprising to find that people's emotions factor in their everyday judgments, nor should it be surprising to find that those emotions often cloud their ability to make good moral judgments. Our history is littered with jealous reprisals and xenophobic discrimination among so many other actions of equal injustice. Some moral rationalists, as I discussed at length in chapter one, argue that we ought not to let our emotions guide our moral judgments. Appealing instead to an inner impartial rational judge is their practical model aimed towards finding a better way of resolving difficult moral dilemmas.

To deny that moral rationalist theories have provided us significant purchase in our attempt to live the good life would be incredible. But those theories also suppose that individuals *can* distinguish between their emotions and reason during rational judgment. If the emotions are characterized as palpable, as bitter disgust or heartfelt affection, it is understandable why emotions and reason have been cast as separate actors in our inner moral lives. Most of us have had first-hand experiences that suggest we *can* control our emotions via reason, sometimes with the difficulty described by Plato, other times with the cool calculation prescribed by Kant, and to a lesser degree by Mill. These experiences make us confident that the emotions and reason *feel* different from one another, thus appearing to support what the moral rationalists suggest, that they can (and ought to) function *disentangled* from one another.

It would be unexpected, therefore, to discover that our emotions play a much more nuanced and complicated role than that supposed by the moral rationalists, even in our most rationally considered moral judgments. How could this be? It could be that the moral rationalist model, which clearly delineates between the influence exerted on our judgments by reason and the emotions, is unwarranted. In other words, it could turn out that the emotions are always implicated in certain moral judgments regardless of our rational intentions. This could be the case if the emotions are not accurately described when framed as the sort of palpable actors as described in the moral rationalist literature. Emotions might more accurately be described as occurring, and being experienced, along a scale, some reactions being palpable and others not, yet all of them somehow implicated in many of our moral rational processes.

In this chapter I will argue that there is good reason to suspect that emotions and reason are not so obviously, or easily, disentangled in moral judgment as the moral rationalist model suggests. I will provide a modest overview of the current empirical research in moral psychology. It calls into question the moral rationalist disentanglement of the emotions and reason in moral judgment, while calling into question the nature of the emotions as they function in moral judgment. I will argue that the empirical data is compelling enough for us to reasonably doubt the model of moral rationalism.

Investigating the Theoretical Separation of Emotions and Moral Reasoning

There is a growing literature in the fields of moral psychology and neurology aimed at exploring the relationship between the emotions and reason in moral judgment.⁹²

Researchers who are contributing to that literature are asking questions about the

⁹² William D. Casebeer, "Moral Cognition and Its Neural Constituents," *Nature Reviews Neuroscience* 4 (October 2003): 841-847.; Joshua Greene, "From Neural 'Is' to Moral 'Ought': What Are the Moral Implications of Neuroscientific Moral Psychology?," *Nature Reviews Neuroscience* 4 (October 2003): 847-850.; Michael Koenigs and Daniel Tranel, "Irrational Economic Decision-Making after Ventromedial Prefrontal Damage: Evidence from the Ultimatum Game," *The Journal of Neuroscience* 27, no. 4 (January 2007): 951-956.; Diana Robertson, et al., "The Neural Processing of Moral Sensitivity to Issues of Justice and Care," *Neuropsychologia* 45 (2007): 755-766.; Jorge Moll, et al., "The Moral Affiliations of Disgust," *Cog Behav Neurol* 18 (2005): 68-78.; Joshua Greene, Leigh E. Nystrom, Andrew D. Engell, John M. Darley and Jonathan D. Cohen, "The Neural Bases of Cognitive Conflict and Control in Moral Judgment," *Neuron* 44 (October 2004): 389-400.; Hauke R. Heekeren, Isabell Wartenburger, Helge Schmidt, Hans-Peter Schwintowski and Arno Villringer, "An fMRI Study of Simple Ethical Decision-Making," *NeuroReport* 14, no. 9 (July 2003): 1215-1219.; Jorge Moll, et al., "The Neural Correlates of Moral Sensitivity: A Functional Magnetic Resonance Imaging Investigation of Basic and Moral Emotions," *The Journal of Neuroscience* 22, no. 7 (April 2002): 2730-2736.; Joshua Greene and Jonathan Haidt, "How (and Where) Does Moral Judgment Work?," *Trends in Cognitive Science* 6, no. 12 (December 2002): 517-523.; Joshua D. Greene, R. Brian Sommerville, Leigh E. Nystrom, John M. Darley and Jonathan D. Cohen, "An fMRI Investigation of Emotional Engagement in Moral Judgment," *Science* 293 (September 2001): 2105-2108.; Marc Hauser, Fiery Cushman, Liane Young, R. Kang-Xin Jin and John Mikhail, "A Dissociation Between Moral Judgments and Justifications," *Mind & Language* 22, no. 1 (February 2007): 1-21.; Michael Koenigs, et al., "Damage to the Prefrontal Cortex Increases Utilitarian Moral Judgments," *Nature* 446 (April 2007): 908-911.; Toshinobu Takeda, Kiyoto Kasai and Nobumasa Kato, "Moral Judgment in High-Functioning Pervasive Developmental Disorders," *Psychiatry and Clinical Neurosciences* 61 (2007): 407-414.; Fiery Cushman, Liane Young and Marc Hauser, "The Role of Conscious Reasoning and Intuition in Moral Judgment," *Psychological Science* 17, no. 12 (2006): 1082-1089.; Mario F. Mendez, Eric Anderson and Jill S. Shapira, "An Investigation of Moral Judgment in Frontotemporal Dementia," *Cog Behav Neurol* 18, no. 4 (December 2005): 193-197.; Shaun Nichols, "Norms With Feeling: Towards a Psychological Account of Moral Judgment," *Cognition* 84 (2002): 221-236.; Jonathan Haidt, Clark McCauley and Paul Rozin, "Individual Differences in Sensitivity to Disgust: A Scale Sampling Seven Domains of Disgust Elicitors," *Person. individ. Diff.* 16, no. 5 (1994): 701-713.

nature of the emotions, particularly whether there is evidence to suggest that they contribute in some way to the overall process of rational moral judgment. As I will demonstrate, the evidence suggests that they do. In addition, it seems that the nature of the emotional contribution to moral judgment does not necessarily square up with the account given by moral rationalists. Emotions seem to function in a far less palpable manner than the winged beasts described by Plato, or the intense sentimental reactions Kant might have anticipated.

The *Is-Ought* Barrier to Empirical Research

Empirical observations about the moral judgments people make often fail to impress moral rationalist philosophers. Facts about what people actually do, they say, have no clear implications on what they ought morally to do, and it is a fallacy to assume otherwise. For example, if it turns out that everyone believes it is acceptable to torture small children, we cannot conclude from that fact that everyone ought to torture small children. In other words *ought* does not follow from *is*. The problem with deriving a moral *ought* from a factual *is* is usually attributed to Hume:

“In every system of morality, which I have hitherto met with, I have always remark'd, that the author proceeds for some time in the ordinary ways of reasoning, and establishes the being of a God, or makes observations concerning human affairs; when all of a sudden I am surpriz'd to find, that instead of the usual copulations of propositions, *is*, and *is not*, I meet with no proposition that is not connected with an *ought*, or an *ought not*. This change is imperceptible; but is however, of the last consequence. For as this *ought*, or *ought not*, that

expresses some new relation or affirmation, 'tis necessary that it shou'd be observ'd and explain'd; and at the same time that a reason should be given; for what seems altogether inconceivable, how this new relation can be a deduction from others, which are entirely different from it."⁹³

Moral rationalists might also say that we commit a similar fallacy, often attributed to Moore and termed the *naturalistic fallacy*, whenever we assume that something natural, that is, something that we tend naturally to do or that is naturally occurring, is morally good.⁹⁴ For example, suppose it were the case that many individuals evolved to treat others violently in order to outperform them. One might make an argument that because we evolved naturally to compete violently it must be the case that competing violently is morally permissible, or morally good. But that would be an instance of the naturalistic fallacy; just because something evolved in such and such a way does not make it morally permissible or good. To counter the example we can point out that we generally place moral restrictions on harming individuals for personal gain, and that we justify those restrictions with reasons. Thus, moral rationalists would say that it is good *reasons* that matter, regardless of what we learn from the natural world. The *is-ought*

⁹³ David Hume, *A Treatise of Human Nature*, ed. L.A. Selby-Bigge and P.H. Nidditch (New York: Oxford University Press, 1978), 469.

⁹⁴ See Chapter 1 of G.E. Moore, *Principia Ethica*, (n.3). Although the *naturalistic fallacy* is often attributed to Moore it is done so somewhat misleadingly. His argument is that the term "good" can be defined neither by metaphysical nor naturalistic terms. In other words, he argued that the term "good" was fundamentally irreducible. The connection between what he terms the naturalistic fallacy and what others attribute to him may stem from his pointing to the problem of defining good in naturalistic terms. Regardless, the naturalistic fallacy tends to take on the same sense as the is-ought distinction (as pointed out by Hume) in the ethics literature.

distinction and the naturalistic fallacy point to the difficulty in incorporating empirical data, which amounts to facts about the way the world *is*, into moral philosophical theories dealing with normative claims, that is, claims about what people *ought* to do.

Hume's *is-ought* distinction can be a formidable problem for anyone interested in incorporating empirical research into theories of moral philosophy. It demands that any normative conclusion that is drawn from empirical research into our moral practices requires some additional justification that would explain why that conclusion has normative moral philosophical import. Again, discovering that everyone believes it is good to do *X* does not in itself justify the claim that people ought to be allowed to do *X*. The facts alone say nothing about what people ought to do.

Philosophers and moral psychologists who are interested in the moral import of empirical facts about moral judgment respond in various ways to the theoretical barrier posed by the *is-ought* distinction. Casebeer, for example, argues that the distinction can be dismissed because it is based on moral philosophical theories that have been defeated in the philosophical literature on other grounds (for example the "collapse" of the analytic/synthetic distinction).⁹⁵ Those theories, he says, "do not stand up against some contemporary naturalized ethical theories"⁹⁶, which have the goal of showing that "norms are natural, and that they arise from and are justified by purely natural processes."⁹⁷ Researchers who support naturalized theories of ethics approach theory building quite differently than moral rationalists such as Kant. Casebeer, for

⁹⁵ William D. Casebeer, "Moral Cognition and Its Neural Constituents," 842 (n.94).

⁹⁶ *Ibid*, 842.

⁹⁷ *Ibid*, 843.

example, considers the correct starting point for the development of a robust moral theory to be facts about the world that guide us in developing an account of morality that squares with our cognitive reality.⁹⁸

Greene, on the other hand, responds to the barrier posed by the *is-ought* distinction with a balanced approach to marrying empirical data with normative moral theories.⁹⁹ He argues that moral naturalists like Casebeer have dismissed the distinction prematurely without any valid justification, and that the distinction must be taken seriously. Though he is skeptical of the kind of naturalized ethics spelled out by Casebeer, he argues that moral naturalists have got one thing right: empirical facts about our moral cognition “can have profound moral implications”, philosophically speaking.¹⁰⁰

Greene considers research into our moral intuitions to have considerable moral import when viewed through an evolutionary lens. Consider two moral dilemmas presented by Unger.¹⁰¹ In the first an individual is driving down a road and comes across an injured hiker with serious bleeding wounds to his legs. The hiker pleads with the driver to take him to the nearest hospital, explaining that he will surely die without the ride. Knowing that the bleeding man will ruin the upholstery of his car—a minor financial inconvenience—the driver opts to ignore the pleas of the hiker, and promptly

⁹⁸ Ibid.

⁹⁹ Joshua Greene, "From Neural 'Is' to Moral 'Ought': What Are the Moral Implications of Neuroscientific Moral Psychology?," (n.94).

¹⁰⁰ Ibid, 847.

¹⁰¹ Peter Unger, *Living High & Letting Die: Our Illusion of Innocence* (New York: Oxford University Press, 1996).

leaves the scene. In the second dilemma the same individual is sitting at home when he receives a letter from a reputable charitable organization. The letter explains that a minor financial donation will save the lives of several individuals in a far off land who are sure to die of a readily preventable disease. The individual ignores the letter and goes about his daily business.

According to a moral rationalist theory like utilitarianism (Unger proffers a utilitarian account of morality) the salient moral features of the hiker dilemma and the envelope dilemma are identical to one another, each is a case in which an individual is asked to suffer a minor financial loss to save a life (several lives in the latter). Utilitarians conclude that the individual ought morally to give a minor financial donation to the aid organization *for the same reasons* that he ought morally to give the hiker a ride. Yet Greene points out that our moral intuitions disagree with that conclusion. Though we would surely condemn the individual for leaving the hiker for dead, we would not judge the dismissal of the letter so harshly; we would likely not be bothered about the dismissal of the letter at all. Moral rationalists who disagree with the utilitarian conclusions drawn by Unger and accept our moral intuitions in this case, Greene argues, are “tempted to assume that there must be ‘some good reason’ why it is monstrous to ignore the needs of someone like the bleeding hiker, but perfectly acceptable to spend our money on unnecessary luxuries while millions starve and die of preventable diseases.”¹⁰²

¹⁰² Greene, "From Neural 'Is' to Moral 'Ought': What Are the Moral Implications of Neuroscientific Moral Psychology?," 849 (n.94). It should be noted that Greene is

Greene considers the evidence of our moral intuitions philosophically insightful in this case. Evolutionary psychology offers an explanation as to why we might judge the two cases differently, one being a situation in which we are confronted up-close-and-personal by someone who needs help, the other a situation where those in need are distant and more abstract actors. Natural selection might explain why the hiker would elicit a strong moral reaction in contrast to the distant individuals since our ancestors would have benefited directly by helping one another when in dire need of assistance.¹⁰³ What the evolutionary theory offers, according to Greene, is an alternative account of why people react the way they do without demanding 'some good reason' from a rationalist perspective. Greene does not argue that there is in fact no good reason to help the hiker over those suffering in far off lands, but he does suggest that the empirical evidence offered up by evolutionary evidence gives us reason to doubt the moral realist project, insofar as rationalists demand some more 'philosophical' reason to explain the difference in our intuitive reactions.

Greene's critique of moral rationalists' tendency to search for reasons rather than accept moral intuitions that are supported by naturalistic explanations, like those offered by Darwinian theories of evolution, does not demonstrate in itself that a tendency to dismiss facts is wrongheaded. He admits as much. However, it draws into focus a problem that moral rationalists face when posed with a conflict between their

specifically arguing against moral realism, as opposed to moral rationalism, in this paper. But the distinction is unimportant to the argument I am making. The point here is that moral rationalists search for reasons to justify the factual intuitions, rather than investigating the normative implications of the intuitions themselves.

¹⁰³ Ibid, 848-849.

rational principles and a seemingly overwhelming moral intuition. If moral rationalists insist on demanding good reasons for accepting the intuition, then it begs the question whether or not those reasons function as anything other than trumped up explanations about factual moral intuitions. If, on the other hand, moral rationalists reject the intuitive pull, settling instead on rational utilitarian or deontological principles, such as the Greatest Happiness Principle or categorical imperative respectively, it still begs the question whether people are justified in holding to their intuitions in dismissing the letter.

Moral rationalists can argue that it does not matter why (or that) people disagree. Simply claiming that individuals ought not to follow their intuitions in the envelope dilemma is a familiar, and accepted, moral rationalist rebuttal to the question of why so many peoples' moral intuitions seem to contradict rational moral principles. Along these lines Singer flatly rejects Greene's attempt to import moral psychological facts into a normative theory of ethics, arguing that evolutionary facts are not relevant in a world so different from the one for which the behavioural traits evolved. In today's world, he says, "We should, with our current powers of reasoning and our rapidly changing circumstances, be able to do better than that."¹⁰⁴

From a moral psychological perspective, Singer's position and any other moral rationalist position that argues against the relevance of empirical data in normative moral theories, ignores the question to what extent individuals are capable of truly settling on such rational principles. To be fair, Singer seems to admit the difficulty of

¹⁰⁴ Peter Singer, "Ethics and Intuitions," 348 (n.78).

pursuing a purely rationally based ethics when he says, “to specify in what sense a moral judgment can have a rational basis is not easy”.¹⁰⁵ Despite this, in his closing remarks he insists on its plausibility: “Nevertheless, it seems to me worth attempting, for it is the only way to avoid moral skepticism.”¹⁰⁶ But if it turns out that empirical data about our moral judgments suggest that there are practical limits to the extent to which we can base our moral judgments on reason, then we have good reason to question the practicalities of the moral rationalist model. This despite any ensuing philosophical worries about moral skepticism, such as those raised by Singer.¹⁰⁷

The theoretical barrier posed by the *is-ought* distinction is not as clear-cut as it seems. Though the barrier is formidable enough to prevent automatic prescriptive moral inferences based solely on facts about the world, it is not necessarily strong enough to prevent more subtle moral inferences to be drawn. Understanding facts about our moral psychology might limit the kinds of claims that can be practically made by moral rationalists who choose to rely solely on theoretical speculation. If their claims include speculations about facts that can be tested empirically, they ought to be tested empirically. Empirical investigation is a relevant and compelling avenue for discovering what, if any, limitations exist.

The Empirical Turn: Moral Psychology and Neurology

The kinds of data produced by Greene and others that can be used to cast doubt on the model of moral rationalism come from empirical investigations into human moral

¹⁰⁵ Ibid.

¹⁰⁶ Ibid.

¹⁰⁷ Ibid.

cognition. These investigations fall under two general categories: functional magnetic resonance imaging (fMRI) studies¹⁰⁸, and more traditional survey type studies¹⁰⁹. In each of these categories the goal of the research is generally to understand what kinds of mental constituents are associated with moral judgment, where judgment in this sense seems to include the deliberative phase of moral problem-solving as I have been using the term. In contrast to the methodology employed by moral philosophers like the rationalists, who describe the various cognitive constituents like reason and emotions using their own mental experiences as a guide, empirical researchers attempt to correlate regions of the brain with particular categories of mental states, such as self-

¹⁰⁸ Michael Koenigs and Daniel Tranel, "Irrational Economic Decision-Making after Ventromedial Prefrontal Damage: Evidence from the Ultimatum Game," (n.94).; Diana Robertson, et al., "The Neural Processing of Moral Sensitivity to Issues of Justice and Care," (n.94).; Jorge Moll, et al., "The Moral Affiliations of Disgust," (n.94).; Joshua Greene, Leigh E. Nystrom, Andrew D. Engell, John M. Darley and Jonathan D. Cohen, "The Neural Bases of Cognitive Conflict and Control in Moral Judgment," (n.94).; Hauke R. Heekeren, Isabell Wartenburger, Helge Schmidt, Hans-Peter Schwintowski and Arno Villringer, "An fMRI Study of Simple Ethical Decision-Making," (n.94).; Casebeer, "Moral Cognition and Its Neural Constituents," (n.94).; Jorge Moll, et al., "The Neural Correlates of Moral Sensitivity: A Functional Magnetic Resonance Imaging Investigation of Basic and Moral Emotions," (n.94).; Joshua Greene and Jonathan Haidt, "How (and Where) Does Moral Judgment Work?," (n.94).; Joshua D. Greene, R. Brian Sommerville, Leigh E. Nystrom, John M. Darley and Jonathan D. Cohen, "An fMRI Investigation of Emotional Engagement in Moral Judgment," (n.94).

¹⁰⁹ Marc Hauser, Fiery Cushman, Liane Young, R. Kang-Xin Jin and John Mikhail, "A Dissociation Between Moral Judgments and Justifications," (n.94).; Michael Koenigs, et al., "Damage to the Prefrontal Cortex Increases Utilitarian Moral Judgments," (n.94).; Toshinobu Takeda, Kiyoto Kasai and Nobumasa Kato, "Moral Judgment in High-Functioning Pervasive Developmental Disorders," (n.94).; Fiery Cushman, Liane Young and Marc Hauser, "The Role of Conscious Reasoning and Intuition in Moral Judgment," (n.94).; Mario F. Mendez, Eric Anderson and Jill S. Shapira, "An Investigation of Moral Judgment in Frontotemporal Dementia," (n.94).; Shaun Nichols, "Norms With Feeling: Towards a Psychological Account of Moral Judgment," (n.94).; Jonathan Haidt, Clark McCauley and Paul Rozin, "Individual Differences in Sensitivity to Disgust: A Scale Sampling Seven Domains of Disgust Elicitors," (n.94).

reference or emotional engagement. For example, the medial orbitofrontal cortex, a region of the brain located (roughly) behind the forehead, is generally implicated in social and emotional behaviour.¹¹⁰ By understanding what regions of the brain seem to correspond to particular kinds of mental states they can try to associate different mental states with particular kinds of cognitive tasks, like moral judgment. If an individual is observed using a region of the brain during some particular moral cognitive task, then particular mental states are implicated in that moral cognitive task. Empirical research into moral judgment is largely an exercise in mapping the parts of the brain that are actively engaged during moral judgment.

In the case of fMRI studies researchers determine which parts of the brain are implicated in moral judgment by producing images detailing the active regions of the brain during moral judgment. Subjects are placed within an fMRI machine and asked to perform moral cognitive tasks, like considering moral dilemmas, listening to descriptions of situations or looking at images depicting moral transgressions. All of these tasks are designed to elicit particular responses correlating to cognitive constituents. For example, Moll, *et al.*¹¹¹ attempted to distinguish between different modes of 'disgust', specifically 'indignation' and 'pure disgust', by asking subjects to read statements like the following:

1. (Pure Disgust) "One night you were walking on a street. You saw a cat eating
its own excrement."

¹¹⁰ Robertson, *et al.*, "The Neural Processing of Moral Sensitivity to Issues of Justice and Care," 762 (n.94).

¹¹¹ Moll, *et al.*, "The Moral Affiliations of Disgust," (n.94).

2. (Indignation) "As you arrived home, you saw that the nurse had put a spider on the baby's face."
3. (Neutral) "You went to the museum and paid for being taught about antiques."¹¹²

Similarly, Robertson, *et al.*¹¹³ asked subjects to consider moral dilemmas some of which were designed to elicit strong emotional responses, others designed to elicit rule-based considerations. Regions of the brain that are 'working' during the experimental observations appear 'activated' in the resulting fMRI image due to an increase in blood flow to those regions. Comparing images of brain activity corresponding to different kinds of moral elicitations allows researchers to draw conclusions about what kinds of cognitive constituents, emotions and the like, are associated with particular kinds of moral engagement.

Survey studies employ a slightly different methodology than those using fMRI imaging technology, though they share the goal of identifying the kinds of cognitive constituents implicated in moral judgment. Mendez, Anderson and Shapira studied subjects with frontotemporal dementia (FTD), a disorder characterized by difficulty in modulating social behaviour.¹¹⁴ According to the researchers FTD patients "lack social

¹¹² Ibid, 71.

¹¹³ Robertson, *et al.*, "The Neural Processing of Moral Sensitivity to Issues of Justice and Care," (n.94).

¹¹⁴ Mendez, Anderson & Shapira, "An Investigation of Moral Judgment in Frontotemporal Dementia," (n.94).

propriety...may perform sociopathic acts [...and] often lack empathy for others” suggesting “alterations in the nature of morality in patients with FTD.”¹¹⁵ One of the “core behavioural” features of FTD is “‘emotional blunting’, a term that includes a loss of empathy or appreciation of the feelings of others.”¹¹⁶ The researchers administered questionnaires designed to gauge empathy and the sense of fairness in the subjects. They also administered two standard moral dilemmas used to elicit emotional responses in the subjects, and asked the subjects to choose what they would do in each case. By interpreting the results of the questionnaires and comparing against the responses of a control group, researchers drew conclusions about the cognitive constituents associated with the frontotemporal regions. They also drew conclusions about the differences in moral tendencies between the control and experimental groups.

The technique employed by Mendez, Anderson and Shapira is common to both types of empirical study attempting to map the moral brain. It involves comparing subjects with previously identified brain injuries, in other words patients displaying abnormal brain function, against ‘normal’ subjects.¹¹⁷ Patients who have suffered physical injury to the brain often display marked changes in behaviour, allowing researchers to infer the kinds of cognitive constituents associated with the damaged

¹¹⁵ Ibid, 193.

¹¹⁶ Ibid.

¹¹⁷ For example Takeda, Kasai & Kato, "Moral Judgment in High-Functioning Pervasive Developmental Disorders," (n.94); Koenigs, *et al.*, "Damage to the Prefrontal Cortex Increases Utilitarian Moral Judgments," (n.94); Mendez, Anderson & Shapira, "An Investigation of Moral Judgment in Frontotemporal Dementia," (n.94).

areas.¹¹⁸ They can also be tested against normal subjects to determine if any differences in cognition are noticeable and can be associated with the particular type of injury, thus allowing them to localize the regions of the brain implicating those kinds of cognitive constituents. Imaging techniques such as fMRI provide researchers with another means of gauging these differences. By comparing the active brain regions of subjects with known cognitive anomalies against normal subjects, regions of the brain can be associated with particular cognitive constituents, such as emotional engagement.

Empirical findings in moral psychology and neurology strongly suggest that the emotions are implicated in all sorts of moral judgments. In a pioneering fMRI study attempting to determine whether emotions were implicated in moral judgment, Greene, *et al.*¹¹⁹ used two classic variations on a moral dilemma, known as the trolley dilemma and footbridge dilemma. A brief restatement of these two dilemmas is in order because they capture so well the intricacies involved in interpreting moral judgments, both from moral philosophical and empirical psychological perspectives.

In the trolley dilemma,

¹¹⁸ Perhaps the earliest example where a localized brain injury was reported as causing specific behavioural changes is the case of Phineas Gage, a rail worker who suffered major brain injury when his tamping rod was blown through his skull, causing a severe localized injury and major changes to his personality. For more information see M. Macmillan, *Phineas Gage Information*, August 28, 2006, <http://www.deakin.edu.au/hmnbs/psychology/gagepage/Pgstory.htm> (accessed May 6, 2008).

¹¹⁹ Greene, *et al.*, "An fMRI Investigation of Emotional Engagement in Moral Judgment,"(n.94).

“A runaway trolley is headed for five people who will be killed if it proceeds on its present course. The only way to save them is to hit a switch that will turn the trolley onto an alternate set of tracks where it will only kill one person instead of five. Ought you to turn the trolley in order to save five people at the expense of one?”¹²⁰

While in the footbridge dilemma,

“[A] trolley threatens to kill five people. You are standing next to a large stranger on a footbridge that spans the tracks, in between the oncoming trolley and the five people. In this scenario, the only way to save the five people is to push this stranger off the bridge, onto the tracks below. He will die if you do this, but his body will stop the trolley from reaching the others. Ought you to save the five others by pushing this stranger to his death?”¹²¹

Most people answer ‘yes’ to the trolley dilemma, and ‘no’ to the footbridge dilemma, a fact that points to a philosophical conundrum.¹²² Why does it seem morally permissible to sacrifice the life of one person to save five in the trolley dilemma, yet morally impermissible to sacrifice the life of one person to save five in the footbridge dilemma?

Greene, *et al.*¹²³ found that areas of the brain associated with emotional response were more active in the footbridge dilemma than in the trolley dilemma, in

¹²⁰ Ibid, 2105.

¹²¹ Ibid.

¹²² Ibid.

¹²³ Ibid.

other words considering the footbridge dilemma more actively engaged our emotions. They describe the idea of pushing an individual to his death as a *moral-personal* dilemma compared with the relatively *moral-impersonal* notion of flipping a switch, despite the identical results. Though the results suggested that personal moral dilemmas were more likely to implicate the emotions, both personal and impersonal moral dilemmas implicated them relative to non-moral dilemmas. Thus, the researchers concluded that emotions are implicated to varying degrees in moral judgment.

Robertson, *et al.*¹²⁴ conducted an fMRI study and proposed that moral sensitivity, that is, an individual's identification of a particular situation as a *moral* situation as opposed to a non-moral or neutral situation, implicates the emotions. In addition, they proposed that moral sensitivity to issues of *care* implicates the emotions to a greater degree than moral sensitivity to issues of *justice*.¹²⁵ Issues of *care* are issues that tend to deal with moral emotions such as empathy and altruism, like deciding to set limitations on work hours to spend time with one's family.¹²⁶ Issues of *justice* are issues that tend to rely more on rule-based processing, like admitting responsibility for an error being blamed on someone else.¹²⁷

¹²⁴ Robertson, *et al.*, "The Neural Processing of Moral Sensitivity to Issues of Justice and Care," (n.94).

¹²⁵ *Ibid.*, 756.

¹²⁶ *Ibid.*

¹²⁷ *Ibid.*

Moll, *et al.*¹²⁸ investigated the differences between *basic* and *moral* emotional responses during moral judgment. Psychologists propose that moral emotions differ from basic emotions “in that they are intrinsically linked to the interests or welfare either of society as a whole or of persons other than the agent.”¹²⁹ Eliciting various emotions in subjects, then having the subjects rate the moral content of the elicitors determines the differences in the emotional responses. Examples of moral elicitors would be “pictures portraying emotionally charged, unpleasant social scenes, representing moral violations (for example, physical assaults, poor children abandoned in the streets, war scenes)”, while basic emotions could be elicited via “unpleasant pictures of aversive scenes not conveying moral connotations (for example, body lesions, dangerous animals, body products).”¹³⁰ By studying fMRI images taken while subjects viewed such situations, they concluded that the emotions are nuanced in that some are more likely to be implicated in moral judgments than others, and that the elicitors matter.

¹²⁸ Moll, *et al.*, "The Neural Correlates of Moral Sensitivity: A Functional Magnetic Resonance Imaging Investigation of Basic and Moral Emotions," (n.94).

¹²⁹ *Ibid.*; Antonio R. Damasio, *Descartes' Error: Emotion, Reason and the Human Brain* (New York: Avon, 1994).; F. de Waal, *Good Natured: the Origins of Right and Wrong in Humans and Other Animals* (Cambridge, MA: Harvard University Press, 1996).; Paul Rozin, Laura Loewry, Sumio Imada and Jonathan Haidt, "The CAD Triad Hypothesis: A Mapping Between Three Moral Emotions (Contempt, Anger, Disgust) and Three Moral Codes (Community, Autonomy, Divinity)," *Journal of Personality and Social Psychology* 76, no. 4 (1999): 574-586.

¹³⁰ Moll, *et al.*, "The Neural Correlates of Moral Sensitivity: A Functional Magnetic Resonance Imaging Investigation of Basic and Moral Emotions," 2731(n.94).

In another study, Moll, *et al.*¹³¹ investigated the range of human disgust responses. According to them, “the diversity of the human disgust experience is grounded in cognitive and emotional mechanisms that give disgust the characteristics of a moral emotion”.¹³² They identify a scale of disgust responses ranging from *basic disgust* to *moral disgust*, described as follows:

“The most general and elementary forms of disgust are *distaste* and *core disgust* that occur in conjunction with offensive and/or aversive sensory experiences. More specific psychologic forms of disgust are typical of humans and occur principally in the social domain, where they are represented by interpersonal and moral disgust. *Interpersonal* disgust is triggered by circumstances or ideas involving people who violate local cultural norms on the use of their bodies, particularly in issues of sex, drugs, and body modification. *Moral* disgust relates to the spiritual protection of the self against degrading and polluting influences.”¹³³

Participants were presented with various cues designed to elicit different types of disgust. Those disgust elicitors were based on the Disgust Scale, proposed by Haidt, McCauley and Rozin, and designed partly to normalize disgust responses across individuals and cultures.¹³⁴ For example, in Moll, *et al.* the elicitors were designed for

¹³¹ Moll, *et al.*, “The Moral Affiliations of Disgust,” (n.94).

¹³² *Ibid*, 68.

¹³³ *Ibid*, 68-69 (my emphasis).

¹³⁴ Jonathan Haidt, Clark McCauley and Paul Rozin, “Individual Differences in Sensitivity to Disgust: A Scale Sampling Seven Domains of Disgust Elicitors,” (n.94).

Brazilian cultural factors, and participants read two sets of statements typified by the following:¹³⁵

1. One night you were walking on a street. You saw a cat eating its own excrement.
2. You went with a friend to a restaurant. When you passed the kitchen you saw rats in the pans.

Participants were asked to rate the statements as evocative of basic disgust or indignation (moral disgust). Results indicated that basic disgust responses are neurologically differentiated from moral disgust responses. This suggests that the disgust emotion is implicated in moral judgments involving characteristics evocative of moral disgust.

Heekeren, Wartenburger, Schmidt, Schwintowski and Villringer¹³⁶ conducted an fMRI investigation to determine if a subset of the regions of the brain associated with moral emotions was implicated in both simple ethical judgments, involving situations devoid of violence and bodily harm, and complex ethical judgments, explicitly evocative of moral emotions. The results of their study suggest that a subset of brain regions associated with moral emotions is indeed implicated in both simple and complex moral judgments.

This data emerging from studies in moral psychology and neurology implicates regions of the brain that are associated with emotional engagement in moral judgment.

¹³⁵ Moll, *et al.*, "The Moral Affiliations of Disgust," 70 (n.93).

¹³⁶ Heekeren, Wartenburger, Schmidt, Schwintowski & Villringer, "An fMRI Study of Simple Ethical Decision-Making," (n.94).

Moral rationalists could argue that discovering the emotions are implicated in moral judgment does not suggest to us that we ought morally to rely on emotions in making moral judgments. We saw this with Singer's response. The force of the *is-ought* distinction bears heavily here, just because subjects in these studies did tangle their emotions up in their more reasonable cognitive processes does not allow the conclusion that they ought to have done so. We ought morally to work towards practically eliminating the role of the emotions in moral judgment through a rigorous adherence to the General Happiness Principle, Categorical Imperative or some other thoroughly rational moral principle.

Moral rationalists may be right in sticking to their principles. But what the data from moral psychology and neurology demonstrate is a fact about certain moral judgments, as they are currently derived. *People's emotions are tangled up in those moral deliberations and judgments.* If we take the moral rationalist objection to the normative relevance of the data seriously, then it seems there exists a gap between the moral rationalist theories and the empirical data. On the one hand, moral rationalists seem to be prescribing the disentanglement of reason and the emotions, while on the other hand the empirical data is suggesting that they are entangled. How can the moral rationalist move from the entangled *is* to the disentangled *ought*?

More precisely, the current empirical data suggests rational-emotional entanglement in moral judgment. If *ought* implies *can*¹³⁷, then we need to empirically investigate the practicalities of the claim that the emotions *can* be disentangled from

¹³⁷ Recall the discussion in Chapter 1 on this implication, which stems from Kant.

reason in moral judgment, or at least to what extent and how. Even if moral rationalists argue that theirs is a weaker claim, that moral judgment ought morally to be premised on *as little emotional input as is rationally possible*, some further investigation seems to be in order.¹³⁸ That weaker claim might be seen as amounting to the prescription that individuals ought not succumb to their emotional reactions. But that clarification does not avoid the issue at hand, it merely restates the presupposition that individuals can disentangle their emotions from their rational faculties during moral judgment. The question remains, to what extent *can* we disentangle emotional from rational input in moral judgment?

Implications For the Model of Moral Rationalism

Data emerging from moral psychology raises significant and challenging questions about the accuracy of the model of moral rationalism, specifically regarding the claim that reason and the emotions can be disentangled in moral judgment. The problem lies partly in the phenomenological nature of the emotions reported in the empirical data. In order to rationally override, or diminish, the role that emotions play in moral judgment one would presumably need to be able to recognize when an emotion is there to diminish. This requires that the emotions feature as phenomenologically apparent components of moral judgment whenever they contribute to the judgment. But it is not clear from the data that the emotions, even though they are often implicated in moral

¹³⁸ Singer, *Ethics and Intuitions*, (n.78) seems to suggest the plausibility of this weaker claim.

judgment, always present themselves as phenomenologically apparent. Long before the advent of fMRI, Hume recognized this possibility:

“’Tis natural for one, that does not examine objects with a strict philosophic eye, to imagine, that those actions of the mind are entirely the same, which produce not a different sensation, and are not immediately distinguishable to the feeling and perception. Reason, for instance, exerts itself without producing any sensible emotion; and except in the more sublime disquisitions of philosophy, or in the frivolous subtleties of the schools, scarce ever conveys any pleasure or uneasiness. Hence it proceeds, that every action of the mind, which operates with the same calmness and tranquillity, is confounded with reason by all those, who judge of things from the first view and appearance. Now ’tis certain, there are certain calm desires and tendencies, which, tho’ they be real passions, produce little emotion in the mind, and are more known by their effects than by the immediate feeling or sensation. These desires are of two kinds; either certain instincts originally implanted in our natures, such as benevolence and resentment, the love of life, and kindness to children; or the general appetite to good, and aversion to evil, consider’d merely as such. When any of these passions are calm, and cause no disorder in the soul, they are very readily taken for the determination of reason, and are suppos’d to proceed from the same faculty, with that, which judges of truth and falshood. Their nature and

principles have been supposed the same, because their sensations are not evidently different.”¹³⁹

Calm passions, hypothesized by Hume, may find a modern correlate in bright spots on an fMRI image indicating activity in regions of the brain commonly associated with emotions. It seems plausible that the calm and thoughtful individual training his thoughts on a moral dilemma might not recognize the activity suggested by the fMRI as an emotional response. It seems there is enough in the hypothesis to cast some doubt on the moral rationalist model, which hinges on phenomenologically apparent emotions in moral judgment. Individuals might be influenced by their emotions without realizing it.

Knowing when the emotions are involved in moral judgment would only satisfy a basic requirement for the moral rationalist. In order to judge properly, one would still need to satisfy several other requirements. First, one would need to provide an explanation as to why it is desirable to disentangle as much as possible the influence of all of the emotions, from the least phenomenologically apparent to the most. Then one would need to overcome the influence of those emotions, which involves the prerequisite of having the ability to do so. Assuming it were possible, overcoming the emotions could prove more difficult than simply deciding to do so, which is the general tone of the moral rationalist model. For instance it might require the application of some targeted conditioning. If so, decisions would need to be made as to the moral permissibility of those methods, including justifications for their application. Choosing

¹³⁹ David Hume, *A Treatise of Human Nature*, 417 (n.95).

to diminish the effect of less phenomenologically apparent emotional responses might demand a different kind of justification than would the choice to diminish the effect of those that are more phenomenologically apparent, such as the ones described by Plato. All of these requirements seem to stem from the recognition that the emotions might not function homogeneously as phenomenologically apparent in our moral deliberations and judgments. We have reason to doubt if the moral rationalist is in the position to argue that the less apparent emotions are the same as the more apparent ones, or that the same arguments and principles apply regarding their dismissal. This seems especially so given the possibility that the data is pointing to a problem with recognizing the contribution that those emotions make to moral judgment, a contribution that might not be so easily overcome.

Data obtained during judgments on the trolley and footbridge dilemmas are helpful in highlighting the problem of phenomenological appearance. Greene, *et al.*¹⁴⁰ report that participants in their study had more difficulty judging the footbridge dilemma, marked by an increase in response time relative to that in the trolley dilemma. They argue that the increase in time was likely caused by the noticeable increase in emotional engagement in the moral-personal dilemma, as observed in the fMRI images. But this does not mean that the subjects *understood* that they were feeling an affective response because of the features of the either dilemma. It could be that their hesitation was only noticeable to the researchers, possibly even only during the

¹⁴⁰ Greene, *et al.*, "An fMRI Investigation of Emotional Engagement in Moral Judgment," (n.94).

post hoc analysis of the data. Even if the participants noticed a relative increase in difficulty judging the footbridge problem, they might not attribute it to an emotional or affective response. It is entirely possible that the participants would attribute the difficulty they experienced in answering the footbridge dilemma to moral rational processes. In other words they might report the relative difficulties being a result of how the dilemmas conflicted with deeply held principles, even deeply held rational principles, but not as a product of any moral emotional engagement prompted by the dilemmas. Yet the data suggest a significant difference in moral emotional engagement between the two judgments.

Data from the studies involving abnormal moral psychologies casts further doubt on the moral rationalist model. Abnormal moral psychologies are characterized in part as impairments to, or as lacking, certain of the emotional cognitive substrates, or components, that are strongly implicated in normal moral judgment. Anderson, Bechara, Damasio, Tranel and Damasio¹⁴¹ report that lesions in regions of the brain associated with moral emotions reduce one's ability to obtain knowledge about standards of moral behaviour and lead to impaired social behaviour and moral decision-making. Their findings, and similar ones, are echoed in studies reporting

¹⁴¹ Steven W. Anderson, Antoine Bechara, Hanna Damasio, Daniel Tranel and Antonio R. Damasio, "Impairment of Social and Moral Behavior Related To Early Damage In Prefrontal Cortex," *Nature Neuroscience* 2, no. 11 (November 1999): 1032-1037.

impaired moral decision-making as a result of damage to those regions of the brain associated with emotion.¹⁴²

In light of the data on abnormal moral judgment, the moral rationalist claim that the role of emotions in moral judgment ought to be disentangled from reason, even the weaker claim that it ought to be minimized, each begs the question: If abnormal moral psychologies are associated with a lack or absence of emotional engagement, that is, a (near) complete disentanglement, then to what extent can we reasonably claim that the emotions ought to be disentangled from the moral equation?

Part of the difficulty in making the rationalist claim in light of the empirical data on moral judgment stems from the relative neurological/psychological complexity of the emotions. The psychological picture painted by moral Rationalists like Plato, Aristotle and Kant is overly simplistic by comparison to the picture emerging from neurology. Our separate emotions do not appear to originate in singular regions of the brain; there is no clumping of cells that is happiness, no neighbouring group of neurons that is disgust. The notion that our individual emotions are like localized beasts each associable with a discrete compartment of the brain is false. And it does not seem likely

¹⁴² Mendez, Anderson & Shapira, "An Investigation of Moral Judgment in Frontotemporal Dementia," (n.94).; J. Grafman, K. Schwab, D. Warden, A. Pridgen, H.R. Brown and A.M. Salazar, "Frontal Lobe Injuries, Violence, and Aggression: A Report of the Vietnam Head Injury Study," *Neurology* 46, no. 5 (1996): 1231-1238.; R.J.R. Blair, "A Cognitive Developmental Approach to Mortality: Investigating the Psychopath," *Cognition* 57 (1995): 1-29.; Antoine Bechara, Antonio R. Damasio, Hanna Damasio and Steven W. Anderson, "Insensitivity to Future Consequences Following Damage to Human Prefrontal Cortex," *Cognition* 50 (1994): 7-15.; Jeffrey L. Saver and Antonio R. Damasio, "Preserved Access and Processing of Social Knowledge In A Patient With Acquired Sociopathy Due To Ventromedial Frontal Damage," *Neuropsychologia* 29 (1991): 1241-1249.

that we can tame them simply by claiming that we must. Rather, the emotions seem to stem from many regions of the brain simultaneously, their intensity and relative cognitive roles similarly regulated by many regions of the brain. Often times the data suggest that a certain part of the brain may be responsible for both the emergence and regulation of a particular range of emotion.¹⁴³ Our understanding of the particulars of these mechanisms is in its infancy. Moral psychologists and neurologists are only starting to separate out functions according to region, and understand the various interactions only at the most basic level. However, there is already enough data available to cast doubt on the simplistic treatment of the emotions provided by moral rationalists.

Conclusion

There are hints in the empirical data emerging from moral psychology and neurology of practical limits to the *ought* claims prescribed by moral rationalist theories. The theoretical barrier posed by the *is-ought* distinction is not formidable enough to prevent relevant moral conclusions to be drawn from facts about the world. Data emerging from moral psychology implicate the emotions in many of our moral judgments. That data also suggests that the influence of the emotions is not so easily separated from the influence of reason in moral judgment. Similarly, the emotions may not be phenomenologically apparent to the individual doing the judging, making the plausibility of the moral rationalist model even more problematic. Indeed, it may be quite difficult to distinguish between a cognitive capacity called 'reason' and another

¹⁴³ Greene & Haidt, "How (and Where) Does Moral Judgment Work?," (n94).

called 'emotion' where moral judgments are concerned. How does one limit the role of the emotions in moral judgment if a) reason and the emotions are tangled up with one another in moral judgment, and b) one is sometimes unaware of the effect that the emotions is having on a moral judgment?

We owe it to the many benefits that moral rationalist theories provided us with to take the data emerging from various sciences at face value, and attempt an understanding of their implications for our moral philosophical theories, especially where those theories intersect our applied ethical practices.

Part of that project includes a call for more empirical research so that we can separate the moral theoretical wheat from the chaff. Insisting that the philosophical alternatives to the current moral rationalist model are problematic, the way that Singer suggests the only alternative is moral skepticism, cannot serve as a reason for dismissing the moral psychological facts. If the facts turn out to limit the central claim of the moral rationalist model, then moral rationalists wishing to developing truly coherent theories must explain how the limits can be incorporated into moral rationalism. We need to develop a language of moral philosophy that squares with the evidence provided by the empirical sciences. It is crucial to note that this is not a call to develop theories that match all of our moral intuitions. The approach that I am suggesting is much more modest. By investigating any psychological limitations to overriding our emotions via reason, we can begin to improve the moral rationalist theories that suppose there are none, if need be.

Anscombe challenged moral philosophers to be wary of moral rationalism until such time that a “sound philosophy of psychology” was at our disposal.¹⁴⁴ She meant to suggest that moral rationalist theories required a conceptual analysis of the terms meant to provide the *moral* character in the prescriptions, terms like “should”, “ought” and “needs”. “For this,” she argued, “we certainly need an account at least of what a human action is at all, and how its description as “doing such-and- such” is affected by its motive and by the intention or intentions in it.”¹⁴⁵ The challenge admits a lack of data in what we now consider the field of moral psychology, and came at a time when little had been unearthed in that field. But if the last decade is an indication of where things are headed, the data from moral psychology appears more sophisticated than that of the past, and it could help philosophers identify empirical claims buried in their theories, while lending them the knowledge required to modify and underpin those theories. We ought to take up Anscombe’s challenge and proceed in trying to stitch together the theories of our inner moral lives with the data that describes it, to whatever extent the data suggests we must.

In Chapter One I argued that moral rationalists presuppose a rudimentary model of moral psychology (which I have called the model of moral rationalism) when they claim that the emotions ought not to play a role in moral judgment. Evidence of that claim can be found in the historical literature, coming from philosophers like Plato, Mill and Kant. Similar claims are echoed in contemporary applied theories of moral

¹⁴⁴ G.E.M. Anscombe, “Modern Moral Philosophy,” 4 (n.74)

¹⁴⁵ *Ibid*, 5.

rationalism, such as the brand of utilitarianism proffered by Singer, and in applied theories of legal adjudication and bioethics. Arguing that the emotions ought not to play a role in moral judgment presupposes that moral agents *can* disentangle reason and emotions in moral judgment, a central claim of the model of moral rationalism. Contemporary moral philosophers have argued that moral rationalism is open to challenges on empirical psychological grounds. Therefore, we have grounds to call the accuracy of the central claim of the moral rationalist model into question until such time that evidence is produced to settle it.

Chapter Two has provided an overview of the current literature emerging from the fields of moral psychology and neurology. Much of the research in those fields is geared towards investigating the relationship between reason and the emotions in moral judgment. That literature strongly suggests that the emotions are implicated in all sorts of moral deliberations and judgments. But the empirical picture that is emerging from moral psychology and neurology gives us reason to question the nature of the emotions in moral judgment. This is especially so when compared to the theoretical account of the emotions provided by moral rationalism. Recent empirical evidence suggests that the influence of the emotions might not be apparent to an individual during moral judgment. This contrasts the historical account, which places the emotions in plain view of the individual during moral deliberation, reducing their dismissal to a matter of rational choice; the individual recognizes that his emotions threaten to interfere with his current moral deliberations and chooses to ignore them, deferring instead to reason as the guiding voice. But if the individual is sometimes unaware that his emotions are actively engaged during moral judgment, then we have

reason to doubt the accuracy and efficacy of the moral rationalist model, which presupposes that awareness as a precursor to the ability to actively disentangle reason and the emotions in moral judgment.

Empirical data from moral psychology and neurology have certainly raised more questions than they have answered. Indeed, the modest goal of my argument is to point out that we have reason to doubt the central claim in the model of moral rationalism, which presupposes that individuals can disentangle reason and the emotions in moral judgment. Testing the accuracy of that claim will undoubtedly require further empirical research. Philosophers will also have to examine the new empirical data, interpret it and incorporate it into their moral philosophical theories accordingly. Ignoring the data and its implications for moral philosophy could amount to a form of philosophical perversity.

If we are to develop an accurate moral philosophy, one consistent with data from the empirical sciences, it will help to focus some of the empirical research on questions specific to theoretical moral philosophy. I have outlined a rudimentary model of moral psychology that is presupposed by moral rationalists. I have also argued that we have reason to question the accuracy of that model. Taken together, these two elements provide the starting point for a more focused approach to an empirically informed moral philosophy aimed at rethinking the model of moral rationalism.

3. Towards A Model of Moral Psychological Multidimensionality

Introduction

In this chapter I will draw on my previous arguments from chapters one and two in order to outline a way forward for empirical research. I will clarify two central questions that must be answered before we can expect to have a fully coherent theory of moral rationalism. Those questions are useful, as they could function as a focal point for further empirical investigation into the moral rationalist model.

I will then suggest an alternative model of moral psychology meant to address the possibility that the moral rationalist model has got it wrong in presuming its central claim. It is based largely on Nichols's *Sentimental Rules* account of moral judgment¹⁴⁶, which is in turn based on work in the empirical sciences, primarily in moral and abnormal psychology. As such, this chapter is rather speculative in spirit. However, the presumptive nature of the current model of moral rationalism makes it speculative in its own right. With this in mind, I will aim to speculate to no greater a degree than the moral rationalists. The alternative model that I will propose could be used as a starting point for a modified theory of moral rationalism, one that accepts the empirical data from moral psychology as a starting point.

¹⁴⁶ Shaun Nichols, *Sentimental Rules: On the Natural Foundations of Moral Judgment* (New York: Oxford University Press, 2004).

Two Questions For Moral Psychology

In the course of this argument two important questions have been raised in light of the empirical data, each of which bears on the accuracy of the model of moral rationalism:

1. To what extent are individuals aware of their moral emotional engagement during moral judgment?
2. To what extent are individuals capable of distinguishing between reason and the emotions in moral judgment?

Question two arises as a direct result of the model of moral rationalism, which contains the central claim that agents are capable of distinguishing between reason and emotion during moral judgment. Empirical data from moral psychology and neurology give us reason to doubt that claim.

A corollary of the second question is that we have reason to question the extent to which we can claim that our emotions and reason are disentangled in the first place. In other words we can ask, "Does the 'rational' aspect of our cognition stand apart from the 'emotional' one, allowing us to objectively observe those 'emotional' aspects while evaluating them on purely 'rational' grounds?" The claim embedded in the moral rationalist model is that we can; it is what Burton refers to as the "myth of the rational mind".¹⁴⁷ Doubting it, as Burton does, leads to the first question.

¹⁴⁷ Robert A. Burton, *On Being Certain: Believing You Are Right Even When You're Not*, 143 (n.72).

Question one challenges the historical philosophical picture of the emotions, which has them appearing like winged beasts and piggish desires, phenomenologically apparent to the individual meant to be morally judging. Data from the empirical sciences shows a complex pattern of physiological activity in areas of the brain associated with moral emotional engagement, during moral judgment. Thus we have reason to question whether or not the emotions that seem to affect our moral judgment are available for us to examine and set aside in the first place. It could be that we are affected in ways that are inaccessible through direct rational examination, no matter how hard we try.

These two questions together highlight a problem with the current normative prescriptions drawn by moral rationalists. As it stands, moral rationalism makes an empirical claim about the nature of moral psychology, expressed in this thesis as the central claim in the model of moral rationalism. At the very least that claim is suspect in light of the current empirical data. At worst it points to incoherence between what moral rationalists claim agents ought to do, and what they are actually capable of doing. Normative moral practices that rely on the moral rationalist model for determining the best judgments in a given situation are similarly threatened. One requirement for a fully coherent rationalist theory must be that it addresses these two questions directly, and in a way that does not contradict the empirical evidence.

The upshot of these two questions is that they provide a focal point for further empirical research into the model of moral rationalism. Studies could be conceptualized to probe the limits of our rational-emotional entanglement, allowing us to better

understand the nature of, and more accurately define, 'reason' and the 'emotions' as aspects of our moral psychology. In the end it may turn out that moral rationalists have essentially got it right, that we can disentangle reason and the emotions during moral judgment, and that we are aware of our moral emotional engagement during those tasks.

At the same time the two questions suggest an alternate possibility: a) that individuals are not always aware of their moral emotional engagement during moral judgment, and b) that individuals are not always capable of completely disentangling reason and the emotions in moral judgment. It might turn out that the moral rationalist model is flawed along these lines. To the extent that moral rationalists have presumed their model, it might prove wise to presume the alternate possibility, to hedge our moral philosophical bets and begin developing an alternate rudimentary model of moral psychology that incorporates these features.

Rethinking the Model of Moral Rationalism

Are the emotions capable of successfully influencing our moral judgments only because moral agents are not trained correctly to dismiss them, as rational agents ought to do? Or are the emotions an irrevocable component of all moral judgments? Or is there some other possibility, some middle ground on which (some) emotions feature as important for certain moral judgments and not for others? Moral rationalists tend to argue in favour of the first of these statements. But if we accept that there is reason to doubt their accompanying psychological presupposition, then other possibilities should be considered.

One recent approach to understanding the implications of empirical research for moral philosophy is found in Nichols's *Sentimental Rules* account of moral judgment.¹⁴⁸ It provides a serious challenge to moral rationalism, one based on empirical evidence from moral psychology. On his account emotions are an essential component of our ability to judge certain social transgressions as moral rather than merely conventional. In developing the Sentimental Rules account Nichols considers two possible models of moral psychology that would support his theory. I will provide an overview of Nichols's main argument and go into some detail regarding his two proposed models of moral psychology. Though I think Nichols's account of moral judgment shares a problematic feature with the moral rationalist model, I also believe it provides an interesting framework upon which I will build a modified model of moral psychology along the lines I have set out.

Core Moral Judgment and Sentimental Rules

Nichols's Sentimental Rules account of moral judgment focuses on a particular class of moral judgments that he terms *core moral judgments*. Those judgments deal with harm-based violations. They are considered at the core of morality because peoples' ability to appreciate them "shows up early ontogenetically" and seems to be "cross-culturally universal".¹⁴⁹ Harm-based violations are those that involve physically harming individuals; a typical example is hitting another person.

¹⁴⁸ Shaun Nichols, *Sentimental Rules: On the Natural Foundations of Moral Judgment* (n.148).

¹⁴⁹ *Ibid*, 7.

Nichols's choice of developing a theory focused on core moral violations is motivated in part by the wealth of literature in moral psychology aimed at understanding how individuals conceptualize moral violations compared with conventional violations.¹⁵⁰ Canonical examples of moral violations used in much of that literature include killing or hitting individuals—they are also examples of core moral violations.¹⁵¹ Conventional, or non-moral, violations tend to vary in the literature but commonly include “violations of school rules, such as not paying attention during story time or talking out of turn.”¹⁵²

Psychologists have developed a robust vocabulary to describe how individuals tend to categorize moral violations as distinct from conventional violations. When research participants are asked to categorize violations, “moral violations attract high ratings on seriousness, they are regarded as having wide applicability, they have a status of authority independence, and they invite different kinds of justifications from conventional violations”¹⁵³ meaning they are considered wrong “because of welfare considerations.”¹⁵⁴ Conventional violations, on the other hand, are considered less serious, more permissible, less generalizably wrong (that is, judgments about them might differ from situation to situation, or culture to culture) and are more authority dependent (that is, they are wrong because some particular authority says so).¹⁵⁵ This

¹⁵⁰ Ibid, 5.

¹⁵¹ Ibid, 6.

¹⁵² Ibid, 5.

¹⁵³ Ibid, 6.

¹⁵⁴ Ibid, 7.

¹⁵⁵ Ibid, 6.

is referred to in the psychology literature as the moral/conventional task. According to Nichols, the findings that support the moral/conventional task are “neither fragile nor superficial” as they have been widely replicated in the scientific literature.¹⁵⁶

Once these kinds of descriptive tasks are well established, researchers can use them to gauge whether or not individuals perform normally on them. Thus, performance on the moral/conventional task can be used as a means of gauging whether or not particular individuals deliver normal moral judgments regarding core moral violations.¹⁵⁷ Consider two individuals, Pat and Terry, each of whom is asked to rate a canonical example of a core moral violation along the dimensions outlined above, say one in which a parent slaps a child hard across the face for speaking out of turn at the dinner table. Suppose Pat rates this violation as highly serious, generalizably wrong and not authority dependent, explaining that it is “horribly wrong to beat an innocent child.” Suppose Terry rates the violation as somewhat serious, moderately generalizable and somewhat authority dependent explaining, “one could get into a lot of serious trouble for beating a child.” Based on the norms developed in the research on core moral judgments, Pat’s performance on the moral/conventional task would be considered normal whereas Terry’s would likely be considered relatively abnormal.

Nichols uses performance on the moral/conventional task as a framework for evaluating moral rationalist claims about the role of emotions in moral judgment. His

¹⁵⁶ Ibid.

¹⁵⁷ I am using the term ‘normal’ simply to indicate a statistical quality, and not to indicate its validity as a normative prescriptive notion. I will discuss general problems associated with this approach later in the chapter.

target is what he calls “empirical rationalism”, which contains a central claim that “it is an empirical fact that moral judgment in humans is a kind of rational judgment; that is, our moral judgments derive from our rational faculties or capacities.”¹⁵⁸ This moral rationalist claim is seriously challenged by empirical evidence in psychology, particularly that surrounding psychopaths. Nichols summarizes the literature that describes psychopaths’ performance on the moral/conventional task:

“Blair compared psychopathic and nonpsychopathic subjects on the moral/conventional task. He found that psychopaths were significantly less likely than nonpsychopaths to treat moral violations as distinctive. Nonpsychopathic criminals, like normal adults and children, made a significant moral/conventional distinction on permissibility, seriousness, and authority contingency; psychopaths, on the other hand, didn’t make a significant moral/conventional distinction on any of these dimensions. More importantly, children with psychopathic tendencies were more likely to judge moral transgressions as authority contingent. Furthermore, psychopaths were less likely than the control criminals to justify rules with reference to the victims’ welfare. Rather, psychopaths typically gave conventional-type justifications for all transgressions...It seems then, that although there is a sense in which

¹⁵⁸ Nichols, *Sentimental Rules: On the Natural Foundations of Moral Judgment*, 67 (n.148). Nichols actually evaluates two separate moral rationalist claims, one that he refers to as the conceptual rationalist claim, and the one stated here, which is the empirical rationalist claim. He argues against both of them on empirical grounds. I will limit my discussion to his treatment of the empirical claim because it encompasses the rationalist model that I have discussed throughout this work. In addition it represents a sizeable body of moral rationalist literature, some of which I have already discussed.

psychopaths do know right from wrong, they don't know (conventional) wrong from (moral) wrong."¹⁵⁹

Despite psychopaths' tendency to perform abnormally on the moral/conventional task it does not seem to stem from a defect in their rational capacities.¹⁶⁰ Psychopaths seem to display the rational capacities necessary for understanding and articulating information about norms/rules, and they demonstrate the ability to rationally apply those norms/rules in distinguishing right from wrong. What they lack is the ability to make the normal moral/conventional distinctions, in other words they tend to treat all violations as conventional in nature. Nor would a defect in their rational capacities necessarily rescue the moral rationalist from the problem posed by psychopaths. Nichols points to evidence from developmental psychology that demonstrates how children display the ability to score normally on the moral/conventional task as early as age three.¹⁶¹ In addition, autistic children and children with Down syndrome, who score low on rational abilities, score normally on the moral/conventional task.¹⁶² All this leads Nichols to conclude that our capacity for core moral judgment does not derive from our rational faculties alone.

The component of core moral judgment overlooked by the moral rationalist, according to Nichols, is an "affective mechanism that is responsive to others' suffering",

¹⁵⁹ Ibid, 76-77.

¹⁶⁰ Ibid, 76.

¹⁶¹ Ibid, 78.

¹⁶² Ibid.

or simply an affective response.¹⁶³ A general sketch of the affective mechanism involves emotions like “contagious distress, personal distress, and concern” that are triggered by the suffering of others.¹⁶⁴ As it turns out, psychopaths exhibit a deficit where these emotions are normally concerned. According to Nichols,

“Nonpsychopathic and psychopathic criminals were shown threatening pictures (for example, an angry face) and pictures of distressed individuals (for example, a crying child). Both Nonpsychopathic and psychopathic criminals show high physiological response to threatening cues. Nonpsychopathic criminals also show a high physiological response to cues of distress in others. By contrast, psychopaths show significantly lower physiological response to distress cues.”¹⁶⁵

Nichols interprets this as an indication that psychopaths are notably characterized by a deficient affective mechanism, which properly accounts for the differences in performance on the moral/conventional task. Thus, in contrast to the moral rationalist, Nichols concludes that core moral judgment can only be accurately explained by appealing to both a “normative theory”, that is, the rational component that sets out rules prohibiting harmful actions, and an affective (emotional) mechanism.¹⁶⁶

We can now set out Nichols’s sketch of the Sentimental Rules account. He argues that core moral judgments derive from a normative theory prohibiting harming others,

¹⁶³ Ibid, 18.

¹⁶⁴ Ibid, 64.

¹⁶⁵ Ibid, 13.

¹⁶⁶ Ibid, 29.

plus an affective mechanism that plays a role in distinguishing moral from conventional violations. However, he leaves open the question of whether or not the affective mechanism is implicated in every moral judgment. More specifically, the lack of current data prevents him from determining whether an affective mechanism is necessary for all core moral judgments, or if an affective mechanism plays an important role in setting up a (minimal) normative moral theory during childhood development, but is not required as a persistent feature for making core moral judgments later in life.

Nichols does, however, sketch out two psychological models meant to represent the possible mechanisms implied by the Sentimental Rules account. The first he calls the *developmental contingency model*, which is premised on the idea “that the affective mechanism and the normative theory are both developmentally necessary for achieving the nonconventional normative theory.”¹⁶⁷ According to this model an individual’s nonconventional normative theory is developed early on in life (probably early childhood) through reinforcement by an affective response mechanism. Repeated experiences could result in the individual treating harm-based violations as more repugnant, which would have the effect of attaching a nonconventional status to rules she develops prohibiting harming others. Whether or not the affective mechanism is necessary after the normative theory is developed is not clear. Nichols suggests that an extreme account of the developmental contingency model would be that after a certain point the affective mechanism simply drops out of the equation, leaving the individual capable of performing normally on the moral/conventional task despite the lack of

¹⁶⁷ Ibid, 26.

affective response. His second model, called the *on-line model*, treats affect as necessary in all core moral judgments.¹⁶⁸ That is, anytime an individual witnesses a core moral violation his normative theory prohibiting harmful actions is complemented by an affective response to establish the nonconventional status of the judgment. An extreme account of the on-line model would have an individual treating purely conventional violations, such as “it’s wrong to put the fork on the right”, as nonconventional, so long as his affective mechanism were triggered during the deliberation.¹⁶⁹ Nichols suggests that neither of these extreme views is likely the case, that alternatively the affective system could play “both a developmental and on-line role in generating and preserving the nonconventional judgments.”¹⁷⁰ In a tone reminiscent of Anscombe’s, Nichols concludes,

“Philosophers [have] feverishly debate[d] whether moral judgment really does depend on the emotions. This philosophical debate has not, to say the least, produced consensus. A more empirically oriented approach might offer greater promise of producing consensus.”¹⁷¹

Nichols’s argument makes it more difficult for the moral rationalist to claim that the emotions and reason ought to be separated during moral judgment. It does so by demonstrating how the emotions might play a crucial role during the developmental stages of one’s moral agency, or into adulthood as an on-line component of moral

¹⁶⁸ Ibid, 28.

¹⁶⁹ Ibid.

¹⁷⁰ Ibid, 29.

¹⁷¹ Ibid.

judgment, or both. His argument suggests that eliminating the emotions altogether might force a new question onto the shoulders of moral rationalists, namely “would such a stark separation between the affective mechanism and reason produce psychopathic-like moral judgments in an otherwise ‘normal’ population?” It is difficult to answer such a question without a greater understanding of the psychological mechanisms at work during moral judgment, but the Sentimental Rules account does seem to make it harder for the moral rationalist to dismiss the empirical findings of moral psychology and neurology and opt simply to reassert that rationalism is the only way to avoid moral skepticism.¹⁷²

Sentimental Rules and the Model of Moral Rationalism

Nichols’s Sentimental Rules account of moral judgment draws from empirical evidence to challenge the claim of moral rationalists that our moral judgments can derive from our rational faculties alone. Data from studies on psychopaths gives reason to believe that rational capacities alone are insufficient to anchor core moral judgment. It also gives reason to believe that the emotions play an essential role in moral judgment. His two psychological models are meant to capture possible mechanisms for explaining how reason and the emotions interact to deliver normal core moral judgments.

I have argued that moral rationalists presuppose a questionable rudimentary model of moral psychology. How does that model compare to Nichols’s two proposed models? Is there room for the moral rationalist to adopt a version of the Sentimental

¹⁷² See CH1 of this thesis, *Singer’s Objection to the Relevance of Moral Psychology* for this discussion.

Rules account while maintaining the presuppositions of the moral rationalist model? And how well does the Sentimental Rules account fare, compared to the model of moral rationalism, regarding the two questions for moral psychology I set out earlier in this chapter? Looking for answers to these questions will help establish whether or not Nichols proposes an account of moral psychology that is compatible with the moral rationalist model, and whether he has avoided the same pitfalls that led us to doubt the moral rationalist model.

Suppose the moral rationalist chooses to accept some very limited role for the emotions in order to try and maintain the presuppositions of the moral rationalist model. This poses a problem, one that is highlighted by using Nichols's psychological models as a guideline to evaluate the moral rationalist's options. Recall that the moral rationalist model has the emotions disentangled from reason during moral judgment, such that moral judgment is not premised on an emotional component. Appealing to the on-line model would, therefore, not be an option for the moral rationalist—the on-line model suggests that emotions are required in every judgment to establish the moral/conventional distinction. Arguing for an extremely diminished on-line affective mechanism would still require the rationalist to admit the emotions into the equation, and that would seem to be exactly what a philosopher like Singer is worried about. Appealing to the developmental contingency model, on the other hand, would allow the moral rationalist to argue that emotions are only important at some early developmental phase of moral agency, and that they play no role in "mature", or fully-rational, agency. Taking this tack commits the moral rationalist to an even more extreme version of the developmental contingency model than that expressed by

Nichols. According to his extreme version, individuals would develop a normative theory with the aid of an affective mechanism early in their moral agency, but would not require the affective mechanism once the normative theory was established. But that version would admit that an individual is capable of making good moral judgments while the affective mechanism is engaged—the agent would simply have no need for the emotions after some critical developmental stage. Indeed, the very point that Nichols is making is that the affective mechanism is what establishes the moral/conventional distinction in the first place, so its active role would not necessarily corrupt the nature of the judgments. Our moral rationalist on the other hand, opting to adopt a developmental contingency model and fit the basic assumptions of the moral rationalist model to it, would have to claim that an individual is only capable of making the best moral judgments once the affective mechanism is disentangled from rational mechanisms. She would have to explain why the judgments the agent was making while developing were corrupted by the emotional component. That would be a problematic argument precisely because the role that the emotional component plays in establishing the moral status of the judgment is considered crucial on the Sentimental Rules account of moral judgment. Thus, corrupted by the emotions, the individual's judgments are not properly formed until the crucial emotional component is dropped from the equation. This is an even more extreme version of the extreme developmental contingency model that Nichols proposes. Thus, it is difficult for the moral rationalist to accept either of the psychological models proposed in the Sentimental Rules account, while maintaining the presuppositions of the model of moral rationalism.

So the Sentimental Rules account poses a serious challenge to moral rationalism based on empirical evidence. It gives reason to doubt the model of moral rationalism by arguing that emotions play a crucial role in moral judgment.

However, Nichols points out an apparent flaw in his account, which he refers to as the *coordination problem*. According to Nichols:

“On the sentimental Rules account, core moral judgment is, of course, tied to the emotions in an important way. There is a nomological connection between having certain emotions and treating certain norms distinctively. What the Sentimental Rules account does not explain, however, is why we happen to have a large cluster of norms that are closely connected to our emotions. We have norms prohibiting harming others, and these norms are closely connected to responses to suffering...But why is it that we have so many norms that fit so well with our emotional repertoire? We might think of this as the *coordination problem*. The Sentimental Rules account has no principled explanation for the coordination between the norms we have and the emotions we have.”¹⁷³

Nichols mitigates this problem by arguing that there are many norms that are not obviously backed by emotions,

“Norms of etiquette (for example the fork is supposed to be on the left)...norms about games (for example do not look at your cards until all the cards have been

¹⁷³ Nichols, *Sentimental Rules: On the Natural Foundations of Moral Judgment*, 115-116 (n.148), Nichols’s emphasis.

dealt); norms about office behavior (for example do not tie up the phone line with personal calls...); norms about school behavior (for example do not chew gum in class...); religious norms (for example go to church on Sunday); and even some putatively moral norms (for example pay your income tax...) are probably not neatly coordinated with emotions.”¹⁷⁴

His concern is that the Sentimental Rules account is merely descriptive, that it provides a “rather accidental” account of certain norms connected with certain emotions (affect in the case of core moral judgment).¹⁷⁵

It may be, however, that Nichols has inadvertently imported the accidental aspect of the Sentimental Rules account, expressed in the form of the coordination problem, from the moral rationalists. Nichols repeatedly refers to the emotions and reason (the normative theory) as “independent”¹⁷⁶, “dissociable”¹⁷⁷ and “different”¹⁷⁸ throughout his argument. His description of the psychological models, indeed much of the Sentimental Rules account, is an attempt to describe “how these mechanisms work together to produce nonconventional normative judgment.”¹⁷⁹ But it is unclear how he justifies the separation between them during moral judgment. His explanation hinges on examples from developmental psychology involving toddler and preschool aged children, and abnormal psychology—the literature on psychopaths—in which

¹⁷⁴ Ibid, 116.

¹⁷⁵ Ibid, 117.

¹⁷⁶ Ibid, 29.

¹⁷⁷ Ibid, 18.

¹⁷⁸ Ibid, 25.

¹⁷⁹ Ibid, 26.

individuals are found to possess one, but not the other, mechanism.¹⁸⁰ But in these cases Nichols admits that the individuals are not making *moral* judgments.¹⁸¹ This is a point I presume rationalists would accept. Why would talk of a separation persist in the discussion of moral judgment?

If one begins to try to understand moral judgment from the position of examining conventional norms, or moral rationalist philosophy, then one is beginning from a standpoint that validates reason but not the emotions. If one then tries to figure out what role the emotions play in moral judgment, without reexamining assumptions, then one runs the risk of trying to figure out how to fit the emotional component into the rationalist picture, as though they demanded some explanation in the first place. Nichols seems to have adopted this methodology, perhaps because the historical moral rationalist tradition is so ubiquitous in the literature. As a result, his models separate reason and the emotions in core moral judgment though his underlying empirical data does not necessarily support that separation. It may be that the separation is an assumption borrowed from the very influential rationalist tradition. Though the *Sentimental Rules* account argues that emotions play some crucial role in distinguishing moral from conventional transgressions, Nichols seems to insist, though to a lesser degree, on the same moral rationalist presupposition regarding the separation of reason and the emotions.

¹⁸⁰ Ibid, 18.

¹⁸¹ Ibid.

So it may be because of Nichols's methodology that the coordination problem appears in the first place. Nichols has set out a group of apparently conventional norms and argued that we need to figure out why other nonconventional norms coordinate with emotions. In the footnote to his argument, Nichols states that the conventional norms he set out (quoted above) might have some underlying emotional component, "a general emotion of disapprobation that attends to all normative violations," but dismisses the probability.¹⁸²

I think his dismissal is premature; his notion of a single general approbation is unlikely, perhaps because it ignores the potential subtleties associated with moral emotional response. Work in ethnography has turned up a surprising list of *human universals*, that is, concepts that appear to be cross-culturally universal.¹⁸³ Some of those, such as "murder proscribed", "healing the sick (or attempting to)", "rape proscribed" and "violence, some forms of proscribed" underscore Nichols's use of affect as a relevant emotion in core moral judgment.¹⁸⁴ Others, such as "cooperation", "conflict, mediation of", "belief in supernatural/religion", "reciprocal exchanges (of labor, goods, or services)" and "statuses and roles" suggest the possibility that we might find evidence beyond rational ability to account for the kinds of norms typically thought to be purely conventional. There is a possibility that many norms, typically thought of

¹⁸² Ibid, 116.

¹⁸³ Steven Pinker, *The Blank Slate: The Modern Denial of Human Nature* (New York: Penguin Books, 2002).

¹⁸⁴ Ibid, 435-439. Pinker lists the universals in an appendix taken from Robert A. Wilson and Frank C. Keil, *The MIT Encyclopedia of the Cognitive Sciences*, ed. Robert A. Wilson and Frank C. Keil (Cambridge, Mass: MIT Press, 1999).

as conventional, may have underlying emotional vestiges owing to our evolutionary past and genetic makeup. To what extent we can disentangle any of those vestiges from our rational processes is as much a question for researchers concerned with conventional norms as it is a question for moral psychologists and philosophers.

This proposition is speculative, but I think Nichols dismisses its possibility prematurely, especially given that it seems to underscore his Sentimental Rules account and simultaneously suggests other avenues for empirical research. If we consider the recent data from moral psychology and neurology I presented in chapter two, we see the emotions taking on a much more nuanced role in moral judgment. It could be that an investigation into other, conventional, norms would generate similar data implicating the emotions. This possibility diminishes the concern Nichols has over the coordination problem, and further raises doubts for the moral rationalist model.

I have suggested an alternative tack to the one taken by moral rationalists and to a lesser degree by Nichols. If we start from the assumption that emotions play a more nuanced, ubiquitous role in moral judgment we might end up with a model of moral judgment that is better suited to testing the limits of the moral rationalist model and moral psychology in general.

A Model of Moral Psychological Multidimensionality

Nichols argues that moral judgments derive from the combination of an emotional response and a rational theory. He identifies an affective mechanism, one that engages when witnessing harm to others, as the emotional response associated with core moral judgment, that is, judgments involving harming others. The hallmark of moral

judgments, when compared to conventional judgments, is that individuals tend to describe them as more serious, less authority dependent, more generalizable and explained by nonconventional language. But the picture he paints is one in which moral and conventional judgments are more or less clearly delineated. Categorizations are determined by querying individuals using a combination of the robust descriptive language of psychology paired with canonical examples of moral violations in the moral/conventional task. In conjunction with this, Nichols draws on evidence from abnormal and developmental psychology as evidence that reason and the emotions both play a role in moral judgment. But children and psychopaths are examples of individuals who are deficient insofar as the capacity for moral judgments is concerned; though children perform normally on the moral/conventional task they are generally not capable of complex reasoning and are not considered morally responsible, whereas criminally convicted psychopaths demonstrate an ability for complex reasoning, but a severe deficiency in affective mechanism. So many clear-cut examples make for a good starting point in examining moral psychology, but it is difficult to see how a more nuanced understanding of the role of emotions in moral judgment might be sought out.

The data suggests we can assume a more nuanced and complex role for the emotions in moral judgment. Canonical examples of moral violations, those designed specifically to elicit strong moral categorizations in the moral/conventional task, seem also to elicit strong emotional responses. What is obscured in the Sentimental Rules account is a description of the variation in performance on the moral/conventional task from individual to individual. Focusing on the variations, and comparing them to the empirical data discussed in chapter two, allows us to develop a rudimentary model of

moral psychology that takes into account the nuanced role that the emotions might play in moral judgment. In addition, drawing from the complexity involved in characterizing abnormal psychologies will help us to understand what kinds of implications a more complex model of moral psychology could have for “normal” moral psychology. I will argue for this model under the rubric of two postulates: (i) moral emotional engagement is gradated; (ii) moral emotional engagement is multidimensional.

Moral Emotional Engagement is Gradated

Data from moral psychology and neurology tend to be summarized to indicate if an experimental group of individuals demonstrated a significantly different moral emotional response when compared to a control group. The use of canonical examples of moral violations designed to elicit strong moral responses is meant to reduce the likelihood of subtle differences in the data, thus improving the statistical results between experimental and control groups. The result is that the analysis obscures variations in the level of emotional response in each individual, both by biasing the emotional response and grouping the results under statistical groupings. The problem with this approach is that it treats the groups as homogeneous in their response.

We can safely assume, though, that variations in the emotional response existed in all of the populations under study in the various examples of research presented throughout this thesis. The variations are implied when researchers report that a certain percentage of subjects respond in such and such a manner. Take a hypothetical example, one reminiscent of the fMRI research presented in chapter two, in which a researcher might claim that 80% of the subjects in an experimental group

demonstrated a moral emotional response to a canonical example of a moral violation. In this hypothetical example, 20% of the subjects' moral emotional engagement was not statistically different from the control subjects, who might have been presented with morally neutral scenarios. Reporting that 80% of the individuals fall within a statistically significant group masks the likely variation within that group by treating them as a homogeneous entity. The same can be said of the 20%.

We would expect a variation in the level, or intensity, of moral emotional engagement from individual to individual given any particular scenario involving a moral violation. The source of the variation is not the issue here—it could be the result of anything from variations in classical conditioning, genetic predispositions or a combination of the two—what is important is that the variation might exist and might influence performance on the moral/conventional task or the kinds of moral judgments an individual might deliver on a particular moral dilemma. Establishing these kinds of influences, and/or the mechanisms explaining them, could be the subject of empirical investigation.

Hints of the influence of variation in moral emotional engagement are evident in the literature already discussed in this thesis. Nichols's evidence from psychopathy gives reason to believe that little or no affective mechanism—the emotional deficit that is one of the hallmarks of the psychopath—influences the psychopath's performance on the moral/conventional task. If Nichols has correctly accounted for the psychopath's performance, compared to that of a nonpsychopathic individual, then the psychopath

might begin to perform differently on the task were his affective mechanism less deficient.¹⁸⁵

Nichols also presents suggestive data regarding the effect of individual disgust thresholds on what he terms the disgusting/conventional task. He reports that particularly disgusting transgressions tend to be judged as more serious, less authority dependent, more generalizable and explained by nonconventional language than less disgusting ones.¹⁸⁶ In addition, individuals who demonstrate low disgust thresholds, that is, they are easily disgusted by particular disgusting transgressions, tend to perform differently on the disgusting/conventional task than high disgust threshold individuals.¹⁸⁷ This variation in performance on the disgusting/conventional task suggests that we might look for a similar effect in individuals with varying affective thresholds.

There is another hint of this evidence provided by Koenigs *et al.*, in which they found that patients with severely diminished moral emotional engagement “produced an abnormally ‘utilitarian’ pattern of judgments on moral dilemmas that that pit compelling considerations of aggregate welfare against highly emotionally aversive behaviours (for example, having to sacrifice one person’s life to save a number of other

¹⁸⁵ I leave it open as to whether or not the developmental contingency or on-line model is the relevant one. I mean only to suggest that a less deficient affective mechanism, at whatever relevant stage in life, would presumably change the psychopath’s performance on the task.

¹⁸⁶ Shaun Nichols, *Sentimental Rules: On the Natural Foundations of Moral Judgment*, 23 (n.148).

¹⁸⁷ *Ibid.*

lives).¹⁸⁸ That is, the patients were more likely to positively endorse utilitarian judgments that involve harming fewer people to save more, where 'normal' patients were less likely to do so. What is suggestive is the overlap between the groups. Perfectly normal utilitarians might deliver the same judgments as individuals with severely impaired moral emotional engagement. This is not meant as a slight against utilitarians (nor against individuals with emotional impairments), rather it is meant to suggest that variations in moral emotional engagement might account for some differences in moral judgment in normal populations. There is currently not enough evidence to fill in the many questions this data raises.

We can begin to sketch out a psychological model that accounts for the variation in moral emotional engagement and the kinds of effects such variation might produce. Nichols argues for a combination of rational and emotional components that somehow produces a nonconventional judgment. Along those lines I would suggest that the level of engagement of the emotional component could play a role in shaping the nature of the judgment in two significant ways. The first is that it could affect performance on the moral/conventional task. Another is that it could affect the nature of the moral judgment, shifting an individual's tendency from one kind of judgment to another for any particular dilemma.

¹⁸⁸ Michael Koenigs, et al., "Damage to the Prefrontal Cortex Increases Utilitarian Moral Judgments," 908 (n.94).

Moral Emotional Engagement is Multidimensional

Nichols focuses on an affective mechanism for explaining core moral judgment, but there could be other emotions that play a role in moral judgment. For example, he describes the role that a disgust mechanism plays in elevating disgusting transgressions along the same descriptive dimensions used in the moral/conventional task. I have already mentioned that disgust has been implicated as a moral emotion by some researchers in the psychology literature.¹⁸⁹ So we have two moral emotions, each of which might play a combinatorial role in conjunction with a normative theory in determining performance on some general nonconventional/conventional task.

Suppose now that we consider the contribution of each moral emotion, as well as the contribution of a normative theory, each as one dimension in a multidimensional moral cognitive space. Suppose also that we postulate the existence of other emotions that play a role in affecting performance on nonconventional/conventional tasks. The overall cognitive system described here can be considered multidimensional in the sense that individual emotional components could be loosely identified that might conspire in complex combinations to account for the variation in judgments.

¹⁸⁹ Jorge Moll, et al., "The Moral Affiliations of Disgust," (n.94).; Jonathan Haidt, Clark McCauley and Paul Rozin, "Individual Differences in Sensitivity to Disgust: A Scale Sampling Seven Domains of Disgust Elicitors," (n.94).

Casebeer, who opts for a more neurobiological basis for his argument, also proposes multidimensionality in moral judgment.¹⁹⁰ His description of multidimensionality is as follows:

“One device that might be useful for helping us to organize abstract moral reasoning is a moral state-space...We can think of much of the activity of frontal cortex and the limbic/brainstem axis as consisting of a moving point in an n -dimensional space, where n could (in complex cases) be determined by making the activity of every neuron that is involved in the system an axis of that space...Reducing the dimensions of that space enables us to capture its principal components, which might themselves correspond to moral concepts that have been explored by ethicists for the last 2,500 years. The idea of a moral state-space allows us to aggregate various cortical regions involved in the processing of moral concepts: if we identify neurons or relevant populations of neurons, and tag each of them as being a dimension of the space...we can reduce the dimensionality of the space to something more manageable.”¹⁹¹

Though Casebeer’s argument seems somewhat far-fetched (he goes on to argue, “the axes of this reduced state-space would correspond to functionally salient groups of neurons, and regions of the state-space might correspond to the ‘big three’ traditional moral theories...or they might help us to identify undiscovered moral concepts”¹⁹²), it shares the basic notion of a complex multidimensional moral cognitive space, and it

¹⁹⁰ William D. Casebeer, "Moral Cognition and Its Neural Constituents," (n.94).

¹⁹¹ Ibid, 845.

¹⁹² Ibid.

seems to suggest a way of explaining utilitarian-biased variations similar to those reported by Koenigs *et al.*¹⁹³

Flanagan also hints at the notion of multidimensionality in moral judgment. He proposes what he calls the Thesis of the Multiple Realizability of Moral Psychologies (TMR), in which he argues that there are infinitely many possible moral psychologies. According to TMR, each moral psychology is largely a product of the particular “social, economic, and institutional arrangements”¹⁹⁴ in which an individual is embedded. Since we have likely not exhausted the complete set of possible manifestations and combinations of these three factors, the set of possible moral psychologies is likely infinite on his account. Though the plasticity of Flanagan’s conception of a moral psychology seems overstated to me (it is not immediately obvious that his characterization of history having “contained a multifarious array of moral personalities”¹⁹⁵ is accurate, nor that we should assume an infinite possible set of social, economic, and institutional arrangements) a constrained version of TMR seems very similar to what I am proposing. A constrained TMR would be the conservative view that there is a smaller set of factors in the form of moral emotions in a multidimensional space, one that exerts a complex influence on human moral psychology. The effects of one’s environment would be something that might be

¹⁹³ Michael Koenigs, et al., “Damage to the Prefrontal Cortex Increases Utilitarian Moral Judgments,” (n.94).

¹⁹⁴ Owen Flanagan, *Varieties of Moral Personality*, 32 (n.75).

¹⁹⁵ *Ibid.*

determined empirically, much like determining the limits of the central claim of moral rationalism.

Thus, I would argue that in addition to moral emotional engagement being gradated, it could be considered multidimensional, with each dimension carrying the potential for gradation. As a start, looking to Brown's list of human universals might provide an interesting touchstone for developing a broad set of dimensions of emotional engagement.

Conclusions

The model of moral psychological multidimensionality that I have sketched out in this chapter has several advantages over the model of moral rationalism. First, it better accounts for the current data emerging from moral psychology and neurology. It does so by presupposing a role for the emotions in moral judgment, similar to that proposed by Nichols. At the same time it improves on the Sentimental Rules account of moral judgment by better accounting for the role of variations in moral emotional engagement in a population. Both Nichols's data and recent data emerging from moral psychology and neurology suggest that variations in the level of moral emotional engagement can influence performance on nonconventional/conventional tasks as well as influencing the kind of moral judgment delivered for a particular moral dilemma. In addition, the model of moral psychological multidimensionality accounts for the potential subtlety and complexity of moral emotional engagement. Both of those aspects are overlooked by the historical moral rationalist account, and to a lesser degree by Nichols, who seems to treat the emotions as somewhat homogeneous and

dissociable, leading to his coordination problem. Yet the subtlety is suggested by the empirical data.

As a touchstone for further empirical research, the model of moral psychological multidimensionality provides a framework from which one might probe the questions posed for moral psychology. An interesting feature of this model is that it accommodates the possibility that the model of moral rationalism might be empirically established. We ought to conduct further research that is specifically aimed at testing the limits of our ability to disentangle our moral emotions from reason while maintaining the semblance of normalcy in our judgments. Without it, moral rationalists will have to settle on presuppositions for the strength of their arguments. That seems unsatisfying, all the more so because we rely on moral rationalist theories to judge on the most difficult dilemmas. Painting an accurate moral psychological picture will go a long way in securing our confidence in whatever judgments turn out to be our best.

4. Postscript - Looking to Future Research

As a natural conclusion to this thesis I had originally planned a lengthy discussion of the particular practical problems associated with the model of moral rationalism. But while researching this work I quickly realized that there is considerable groundwork to be laid in preparation for that discussion. Without the groundwork the discussion would be almost purely speculative. My research prompted me to begin laying that groundwork, which I think I have done with some success. There is a historical trend among moral rationalists, which includes a particular characterization of human emotions. That characterization involves a problematic presupposition about our moral psychology, one that does not seem to be supported by current empirical research into moral psychology and neurology. We can pose specific questions to moral rationalists about this problematic disjoint between how they argue we ought morally to judge, and how it seems we judge, as a matter of fact. Asking those questions does not open us up to challenges on the grounds of a distinction between what *is* and what *ought* to be. Without more research there is little hope of settling the questions. So beyond making the case that we have reason to doubt the model of moral rationalism this thesis has tread on speculative grounds, set out in chapter three.

Drawing practical conclusions on speculative grounds, at this point, would be an exercise in heaping speculation upon speculation. In my estimation it would detract from the primary argument I have chosen to settle on. I think this mine was the prudent approach; until there is some evidence to fill in the blanks of a coherent model of moral psychology, even the moral rationalists are stuck unable to make definitive

prescriptions. Finding the missing pieces is a daunting project, certainly one larger than the scope of a Master's thesis. But I think it is one worth pursuing, if only to help avoid some practical pitfalls we may already find ourselves in.

Despite my reluctance to commit to further speculation, I do sense that future evidence will not be in favour of the model of moral rationalism. The trend in the current research suggests that: (i) moral emotions are often entangled with reason in moral judgment, (ii) moral emotional engagement is gradated, and (iii) moral emotional engagement is multidimensional. These kinds of findings make it harder to imagine how we could argue for a strong disentanglement between moral emotions and reason. They also make it harder to imagine how we might argue for moral objectivism along moral rationalist lines, that is, it seems less and less likely that reasons will provide the objective anchors for morality if those reasons are subject to entanglement with our moral emotions—so much the worse for the moral rationalist model of moral psychology.

I would like to offer a glimpse of what I think is one of the most interesting practical questions—indeed one of the potential pitfalls I just eluded to—that might be researched in the event that the moral rationalist model is undermined by future evidence. If our moral psychology turns out to be multidimensional, gradated, and even weakly entangled, as I have proposed, then it seems that any manipulation of our moral emotional states could have a profound impact on our moral judgments. What does this say about methods we develop that purposely or inadvertently affect the emotions?

There are many examples of targeted manipulations of this sort. On the more commonplace end of the scale paint colours have long been known to affect individuals' moods. Hospital rooms are generally painted in light shades of green or yellow. Caffeine has long been linked to anxiety and irritability, but has more recently been researched for its nootropic effects, resulting in its additive in so-called "smart" drinks. At the more speculative and exotic end of the scale, current offerings of neuropharmaceuticals designed to make people happier, more aggressive, less shy, more alert and less depressed all act on what we would currently understand as emotions.

The moral rationalist model protects our good moral judgments from undue influence from these kinds of manipulations by arguing that reason is disentangled from the emotions. Therefore, on the moral rationalist model, manipulating the moral emotions should have no impact on one's moral character, so long as one is prepared to stick to his rationalist principles; moral character could only be affected by taking a pill that made one less rational. But if the moral rationalist model collapses along the lines I have suggested, then any manipulation of a moral emotional state could potentially affect moral judgment. This is not to say that we could all accidentally become psychopaths, though that outcome might be predicted depending on the nature and degree of the manipulation. Rather, it is the effect of subtler manipulations that could have the most unexpected consequences on our judgments. Individuals might tend more or less towards consequentialist judgments than they would have prior to the manipulation, a possibility foreshadowed by some of today's empirical evidence. Those kinds of subtle changes, if not subject to correction via strict adherence to moral

rationalist principles alone, could have a profound impact on one's personality, and consequently their identity.

Of course, the degree of speculation here is large, so I think it prudent to step away from the speculation and to merely suggest that there are many unanswered questions that could be researched as a result of the groundwork laid in this thesis, each of which holds implications for our theories in philosophy.

5. Bibliography

Miriam-Webster. 1982. [http://www.merriam-](http://www.merriam-webster.com/dictionary/jury+nullification)

[webster.com/dictionary/jury+nullification](http://www.merriam-webster.com/dictionary/jury+nullification) (accessed July 29, 2008).

Anderson, Steven W., Antoine Bechara, Hanna Damasio, Daniel Tranel, and Antonio R. Damasio. "Impairment of Social and Moral Behavior Related To Early Damage In Prefrontal Cortex." *Nature Neuroscience* 2, no. 11 (November 1999): 1032-1037.

Anscombe, G.E.M. "Modern Moral Philosophy." *Philosophy* 33, no. 124 (January 1958): 1-19.

Aristotle. "Nicomachean Ethics." In *Nicomachean Ethics*, translated by Sarah Broadie and Christopher Rowe, 95-258. Oxford: Oxford University Press, 2002.

Bechara, Antoine, Antonio R. Damasio, Hanna Damasio, and Steven W. Anderson.

"Insensitivity to Future Consequences Following Damage to Human Prefrontal Cortex." *Cognition* 50 (1994): 7-15.

Blair, R.J.R. "A Cognitive Developmental Approach to Mortality: Investigating the Psychopath." *Cognition* 57 (1995): 1-29.

Broadie, Sarah. "Philosophical Introduction." In *Nicomachean Ethics*, by Aristotle, edited by Sarah Broadie and Christopher Rowe, 9-91. Oxford: Oxford University Press, 2002.

Burton, Robert A. *On Being Certain: Believeing You Are Right Even When You're Not*. New York: St. Martin's Press, 2008.

- Casebeer, William D. "Moral Cognition and Its Neural Constituents." *Nature Reviews Neuroscience* 4 (October 2003): 841-847.
- Copleston, Frederick S.J. *A History of Philosophy: Modern Philosophy, Part II, Kant*. Vol. VI. Garden City, N.Y.: Image Books, 1960.
- Cushman, Fiery, Liane Young, and Marc Hauser. "The Role of Conscious Reasoning and Intuition in Moral Judgment." *Psychological Science* 17, no. 12 (2006): 1082-1089.
- Damasio, Antonio R. *Descartes' Error: Emotion, Reason and the Human Brain*. New York: Avon, 1994.
- de Waal, F. *Good Natured: the Origins of Right and Wrong in Humans and Other Animals*. Cambridge, MA: Harvard University Press, 1996.
- Denise, Theodore C., Nicholas P. White, and Sheldon P. Peterfreund. "Duty and Reason." In *Great Traditions in Ethics*, by Theodore C. Denise, Nicholas P. White and Sheldon P. Peterfreund, edited by Theodore C. Denise, Nicholas P. White and Sheldon P. Peterfreund, 143-156. Belmont, CA: Wadsworth, 2002.
- Finnis, John. "Natural Law and Legal Reasoning." In *Law and Morality: Readings in Legal Philosophy*, edited by David Dyzenhaus and Arthur Ripstein, 200-217. Toronto: University of Toronto Press, 2001.
- Flanagan, Owen. *Varieties of Moral Personality*. Cambridge, Mass.: Harvard University Press, 1991.
- Glannon, Walter. *Biomedical Ethics*. New York: Oxford University Press, 2005.

Grafman, J., K. Schwab, D. Warden, A. Pridgen, H.R. Brown, and A.M. Salazar. "Frontal Lobe Injuries, Violence, and Aggression: A Report of the Vietnam Head Injury Study." *Neurology* 46, no. 5 (1996): 1231-1238.

Greene, Joshua D., R. Brian Sommerville, Leigh E. Nystrom, John M. Darley, and Jonathan D. Cohen. "An fMRI Investigation of Emotional Engagement in Moral Judgment." *Science* 293 (September 2001): 2105-2108.

Greene, Joshua. "From Neural 'Is' to Moral 'Ought': What Are the Moral Implications of Neuroscientific Moral Psychology?" *Nature Reviews Neuroscience* 4 (October 2003): 847-850.

Greene, Joshua, and Jonathan Haidt. "How (and Where) Does Moral Judgment Work?" *Trends in Cognitive Science* 6, no. 12 (December 2002): 517-523.

Greene, Joshua, Leigh E. Nystrom, Andrew D. Engell, John M. Darley, and Jonathan D. Cohen. "The Neural Bases of Cognitive Conflict and Control in Moral Judgment." *Neuron* 44 (October 2004): 389-400.

Haidt, J. "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment." *Psychological Review* 108, no. 4 (2001): 814-834.

Haidt, Jonathan, Clark McCauley, and Paul Rozin. "Individual Differences in Sensitivity to Disgust: A Scale Sampling Seven Domains of Disgust Elicitors." *Person. individ. Diff.* 16, no. 5 (1994): 701-713.

Hauser, Marc, Fiery Cushman, Liane Young, R. Kang-Xin Jin, and John Mikhail. "A Dissociation Between Moral Judgments and Justifications." *Mind & Language* 22, no. 1 (February 2007): 1-21.

Heekeren, Hauke R., Isabell Wartenburger, Helge Schmidt, Hans-Peter Schwintowski, and Arno Villringer. "An fMRI Study of Simple Ethical Decision-Making." *NeuroReport* 14, no. 9 (July 2003): 1215-1219.

Hume, David. *A Treatise of Human Nature*. Edited by L.A. Selby-Bigge and P.H. Nidditch. New York: Oxford University Press, 1978.

Kant, Immanuel. *Critique of Practical Reason*. Edited by Mary Gregor. Translated by Mary Gregor. Cambridge: Cambridge University Press, 1997.

Kant, Immanuel. "On a Supposed Right to Lie because of Philanthropic Concerns." In *Grounding for the Metaphysics of Morals*, by Immanuel Kant, translated by James Wesley Ellington. Hackett Publishing, 1993.

Kant, Immanuel. "The Foundations of the Metaphysics of Morals." In *Ethical Theory: A Concise Anthology*, edited by Heimer Geirsson and Margaret R. Holmgren, 113-130. Peterborough, Ontario: Broadview Press, 2000.

Kluge, Eike-Heiner W. *Readings in Biomedical Ethics: A Canadian Focus*. 3rd Edition. Edited by Eike-Heinier W. Kluge. Toronto: Pearson Prentice Hall, 2005.

Koenigs, Michael, and Daniel Tranel. "Irrational Economic Decision-Making after Ventromedial Prefrontal Damage: Evidence from the Ultimatum Game." *The Journal of Neuroscience* 27, no. 4 (January 2007): 951-956.

Koenigs, Michael, et al. "Damage to the Prefrontal Cortex Increases Utilitarian Moral Judgments." *Nature* 446 (April 2007): 908-911.

"List of CCAC Guidelines." *Canadian Council on Animal Care*. February 22, 2008.
http://www.ccac.ca/en/CCAC_Programs/Guidelines_Policies/GDLINES/Guidelis.htm
(accessed August 28, 2008).

Macmillan, M. *Phineas Gage Information*. August 28, 2006.
<http://www.deakin.edu.au/hmnbs/psychology/gagepage/Pgstory.htm> (accessed May 6, 2008).

Mendez, Mario F., Eric Anderson, and Jill S. Shapira. "An Investigation of Moral Judgment in Frontotemporal Dementia." *Cog Behav Neurol* 18, no. 4 (December 2005): 193-197.

Mill, John Stuart. "Utilitarianism." In *Ethical Theory: A Concise Anthology*, edited by Heimer Geirsson and Margaret R. Holmgren, 86-94. Peterborough, Ontario: Broadview Press, 2000.

Moll, Jorge, et al. "The Moral Affiliations of Disgust." *Cog Behav Neurol* 18 (2005): 68-78.

Moll, Jorge, et al. "The Neural Correlates of Moral Sensitivity: A Functional Magnetic Resonance Imaging Investigation of Basic and Moral Emotions." *The Journal of Neuroscience* 22, no. 7 (April 2002): 2730-2736.

Moore, G.E. *Principia Ethica*. Cambridge: Cambridge University Press, 1965.

National Council on Ethics in Human Research. "Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans." *National Council on Ethics in Human Research*. 1999. http://www.ncehr-cnerh.org/english/code_2/ (accessed June 16, 2008).

Nichols, Shaun. "Norms With Feeling: Towards a Psychological Account of Moral Judgment." *Cognition* 84 (2002): 221-236.

—. *Sentimental Rules: On the Natural Foundations of Moral Judgment*. New York: Oxford University Press, 2004.

Pinker, Steven. *The Blank Slate: The Modern Denial of Human Nature*. New York: Penguin Books, 2002.

Plato. *Phaedrus*. Translated by Christopher Rowe. New York, New York: Penguin Books, 2005.

—. *Republic*. Translated by G.M.A. Grube and C. D.C. Reeve. Indianapolis, Indiana: Hackett Publishing Company, 1992.

R. v. Krieger. [2006] 2 S.C.R. 501, 2006 SCC 47

R. v. Latimer. [2001] 1 S.C.R. 3, 2001 SCC 1

R. v. Morgentaler. [1988] 1 S.C.R. 30

Robertson, Diana, et al. "The Neural Processing of Moral Sensitivity to Issues of Justice and Care." *Neuropsychologia* 45 (2007): 755-766.

Rozin, Paul, Laura Loewry, Sumio Imada, and Jonathan Haidt. "The CAD Triad Hypothesis: A Mapping Between Three Moral Emotions (Contempt, Anger, Disgust) and Three Moral Codes (Community, Autonomy, Divinity)." *Journal of Personality and Social Psychology* 76, no. 4 (1999): 574-586.

Saver, Jeffrey L., and Antonio R. Damasio. "Preserved Access and Processing of Social Knowledge In A Patient With Acquired Sociopathy Due To Ventromedial Frontal Damage." *Neuropsychologia* 29 (1991): 1241-1249.

Singer, Peter. "Ethics and Intuitions." *The Journal of Ethics* 9, no. 3/4 (October 2005): 331-352.

Sneddon, Andrew. "Feeling Utilitarian." *Utilitas* 15, no. 3 (November 2003): 330-352.

Stern, Robert. "Does 'Ought' Imply 'Can'? And Did Kant Think It Does?" *Utilitas* (Cambridge University Press) 16, no. 1 (March 2004): 42-61.

Takeda, Toshinobu, Kiyoto Kasai, and Nobumasa Kato. "Moral Judgment in High-Functioning Pervasive Developmental Disorders." *Psychiatry and Clinical Neurosciences* 61 (2007): 407-414.

Taylor, Richard. *Good and Evil: A New Direction*. London: Collier-Macmillan Ltd., 1970.

Unger, Peter. *Living High & Letting Die: Our Illusion of Innocence*. New York: Oxford University Press, 1996.

Wilson, Robert A., and Frank C. Keil, . *The MIT Encyclopedia of the Cognitive Sciences*. Cambridge, Mass: MIT Press, 1999.