

NATURALIZING MORAL JUDGMENT

By

Theresa K. Pecoskie

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Arts
in
Philosophy

Waterloo, Ontario, Canada, 2006

© Theresa K. Pecoskie 2006

AUTHOR'S DECLARATION FOR ELECTRONIC SUBMISSION OF A THESIS

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners. I understand that my thesis may be made electronically available to the public.

NATURALIZING MORAL JUDGMENT

ABSTRACT

Philosophers have traditionally attempted to solve metaethical disputes about the nature of moral judgment through reasoned argument alone. Empirical evidence about how we *do* make moral judgments is often overlooked in these debates. In the wake of recent discoveries in cognitive neuroscience and experimental psychology, however, some empirically-minded philosophers are beginning to use neural findings in support of their theories of moral judgment. The intent of this thesis is to explore how this empirical evidence can be integrated effectively into philosophical discussions about moral judgment. In the first chapter of my thesis, I review the moral judgment debate in both philosophy and moral psychology, focusing specifically on contemporary sentimentalist solutions to this problem. This review sets the stage for my critique of Prinz's sentimentalist account of moral judgment in the second chapter. I argue that Prinz uses neural evidence to support his sentimentalist thesis inappropriately, altering the evidence to fit his theory, rather than using the evidence to inform his theory. In the third chapter, I examine Prinz's somatic theory of emotion and how this is related to his theory of moral judgment. I argue that neural evidence indicates that a theory of emotion that incorporates aspects of both cognitive appraisal and somatic theories is more empirically accurate than either view in isolation. Finally, I discuss the implications that a neural account of emotion could have on future debates about the nature of moral judgment.

ACKNOWLEDGEMENTS

First, I would like to thank Paul Thagard for his patience and encouragement. Thank you also for opening my eyes to a new way of understanding old philosophical problems.

Second, I would like to thank Debbie Dietrich and Linda Daniel for helping me stay organized and on track while completing this project.

Finally, I would like to thank my family and friends who have been there for me through all of the triumphs and frustrations of the past year. I could not have finished this project without your constant love and support.

TABLE OF CONTENTS

Introduction.....	1
Chapter 1: Metaethical Views of Moral Judgment.....	4
Chapter 2: Prinz’s Sentimentalist View of Moral Judgment.....	21
Chapter 3: Implications for Future Theories of Moral Judgment.....	41
Conclusion.....	57
References.....	59

TABLES AND ILLUSTRATIONS

Figure 1: Theories of Moral Judgment.....	5
Figure 2: The EMOCON Model of Emotional Consciousness.....	52

Introduction

Philosophers have traditionally tried to solve metaethical disputes about the nature of moral judgment through reasoned argument alone. Empirical evidence about how we *do* make moral judgments has typically been overlooked in these debates. Recently, however, some important discoveries have been made in cognitive neuroscience about the neural underpinnings of moral judgment. Many empirically-minded philosophers are starting to take these findings into consideration when formulating their arguments. The general intent of this thesis is to explore how neuroscientific information is being integrated into theories of moral judgment, to evaluate how effective these findings are in supporting theories of moral judgment, and to explain how neuroscience has the power to impact future metaethical debates about the nature of moral judgment.

I will open my thesis with a brief, but necessary explanation of the historical metaethical debate over moral judgment in both philosophy and moral psychology. On one side of the debate are the rationalists, who argue that our judgments are ultimately grounded in reason. Challenging this rationalist claim are the sentimentalists, who consider our judgments to be fundamentally emotional. After surveying the history of the debate, I will move forward to describe how this debate is playing out amongst sentimentalists in contemporary metaethical circles. There is presently a deep divide within sentimentalism itself over the nature of the relationship between our emotions and our judgments. I will discuss both the cognitivist view that our moral judgments express beliefs about our sentiments as well as the non-cognitivist view that our moral judgments function simply to express our sentiments. I will then move forward to discuss the

“conflation problem” (D’Arms and Jacobson 2000), which is an interesting difficulty that many contemporary sentimentalists, cognitivist and non-cognitivist alike, are faced with.

Explaining the history of the moral judgment debate is needed to set the stage for my examination of Jesse Prinz’s (2006) sentimentalist theory of moral judgment in the second chapter of my thesis. Prinz’s theory has been selected because it exemplifies a growing trend in metaethical research. This trend is to integrate empirical evidence from experimental psychology and cognitive neuroscience with one’s theoretical intuitions in order to present a strong, empirically plausible theory. After giving a brief overview of Prinz’s position, I will argue that the empirical data he cites is incapable of supporting his sentimentalist view. I will argue further that Prinz’s view nonetheless reveals some important clues as to how empirical data can effectively be integrated into future discussions about moral judgment.

In my final chapter, I will argue that Prinz’s sentimentalist theory does not fail merely because he uses empirical evidence ineffectively, but also as result of the view of emotion that underlies his theory. I will examine the finer points of Prinz’s somatic view of emotion before moving on to discuss cognitive appraisal theories of emotion and somatic theories of emotion more generally. I will argue that theorists on both sides of the moral judgment debate have misconstrued what emotion is. I will then propose that a theory of emotional consciousness that integrates elements from cognitive appraisal theories with elements from somatic theories is a more neurally plausible alternative than either theory in isolation.

My paper will conclude with a discussion of how an empirically informed, integrative theory of emotion could influence future metaethical debates about the nature

of moral judgment. My discussion will focus on the idea that, when used effectively, empirical evidence from cognitive neuroscience and experimental psychology has the potential to drive philosophical discussions of moral judgment forward. Progress, however, will be made at the expense of giving up some of our more dogmatic philosophical beliefs about reason, emotion, and moral judgment in general.

Chapter 1: Metaethical Views of Moral Judgment

Before I discuss Prinz's view in more detail, it will be helpful if I provide an outline of the metaethical debate that Prinz is entering into. I have included a chart that may make reading through the often confusing metaethical material slightly easier (see Figure 1 below). I will open this section with an overview of the traditional debate in both philosophy and moral psychology over the nature of moral judgment. On one side of this debate are the rationalists, who hold that we arrive at our judgments through reason alone. Sentimentalists, on the other hand, argue that our judgments are fundamentally emotional. Next, I will move forward to discuss the current debate between the sentimentalists who endorse a cognitivist view of moral judgment, such as David Wiggins and John McDowell, and those sentimentalists who argue in favour of non-cognitivism, such as Allan Gibbard and Simon Blackburn. Although the views of Wiggins and McDowell are at odds with those of Gibbard and Blackburn, I argue that all of these theories are plagued by a common problem. According to D'Arms and Jacobson (2000), these contemporary sentimentalist theories are all vulnerable to the "conflation problem." I will conclude this section by explaining what the conflation problem amounts to and by detailing how Prinz's sentimentalist account of moral judgment is directed specifically at avoiding this neo-sentimentalist difficulty.

1.1 Moral Judgment in Philosophy

The traditional debate in philosophy over the nature of moral judgment is between the rationalists and the sentimentalists. Moral rationalism stresses, in some way or another, our ability to grasp moral principles through a priori reason. Though the moral principle in question is often considered objective by rationalists, this need not be the case. Moral

Figure 1: Theories of Moral Judgment

	Rationalism Moral judgments are arrived at through reason alone.	Sentimentalism Moral judgments are fundamentally emotional.
Cognitivism Moral judgments express beliefs.	Plato Kant Railton Mackie Piaget Kohlberg	Wiggins McDowell Prinz Gilligan Mill
Non-Cognitivism Moral judgments do not express beliefs.		Epicurus Hume Smith Ayer Blackburn Gibbard

judgments such as “X is bad” or “X is wrong,” therefore, result when subjects deem “X” to be wrong through the use of reason or introspection alone.

Consider, for example, the Platonic view. Plato (1991) contends that the human soul is composed of three separate parts: reason, emotion, and appetite. In the majority of individuals, emotion or appetite will rule their character. Their moral judgments and decision-making will be guided by emotional and appetitive drives and they will aim at fulfilling their immediate desires. In a minority of cases, however, reason will exert control over emotion and desire. Only then will the individual be rendered capable of recognizing objective moral values (i.e., the Form of the Good) and consequently making moral judgments guided by these objective reasons.

On the Platonic view, reason and emotion are taken to be distinct and conflicting human faculties. Like a charioteer and his horses, reason must reign in and rule emotion by whatever means necessary (Plato 2005). It is only when reason exerts control over this faculty that the soul is in harmony and the individual is capable of making appropriate

moral judgments. Emotion impedes the ability to recognize and act in accord with moral values, if it is not overruled.

If Plato exemplifies classical rationalism, then Epicurus should be considered to exemplify classical sentimentalism. Sentimentalism, recall, views moral judgments as fundamentally emotional. Judgments like “X is wrong” or “X is bad” are not arrived at through reason or an a priori grasp of some objective moral norm, but rather, through a negative emotional reaction to “X”. Hence, “X” is bad only insofar as it elicits a negative emotional response from an observer. Epicurus affirms this, arguing that sensation and emotion are the only standards by which we are able to make proper judgments (Diogenes Laertius 1925). On the Epicurean account, something is good if it elicits pleasure in the body or soul and bad if it elicits pain. Pleasure, says Epicurus, “is our first and kindred good. It is the starting point of every choice and of every aversion, and to it we come back, inasmuch as we make feeling the rule by which to judge of every good thing” (Diogenes Laertius 1925, p. 655).

Epicurus additionally argues, contrary to Plato, that reason and emotion, though distinct, need not be opposed to one another. Instead, emotion and sensation can help to inform our beliefs and choices. They direct us towards reasonable judgments, rather than away from them. Emotion, moreover, provides a standard by which we can evaluate our judgments post hoc. Any moral judgment that, upon reflection, produces pleasure and prevents pain is admirable (Diogenes Laertius 1925).

Plato’s rationalist views had a profound influence on the Stoics, Medieval thinkers such as Anselm, Augustine, and Aquinas, and 17th century continental philosophers such as Leibniz and Descartes. Epicurus’ sentimentalism inspired writers

such as Diogenes Laertius and Lucretius in the 3rd century CE and the 1st century BCE respectively. Throughout the Medieval period, Epicurean sentimentalism lost popularity because this materialist view conflicted with popular Christian doctrine. The view was revived, however, by Gassendi in the 17th century and later by 18th century Scottish philosophers such as Hume and Smith. Hume and Smith, like their Epicurean predecessors, argued that judging something morally wrong depends on the presence of a sentiment of disapproval towards that thing. Thus to judge, for instance, that murder is wrong, is simply to have a sentiment of disapproval (e.g., guilt, remorse, anger, and so forth) towards murder.

Hume agrees with the ethical writers before him that reason and emotion are distinct capacities. He does not, however, agree with the rationalist view that they conflict with one another. Only an emotion can conflict with a contrary emotion, not a belief (Hume 1967). Moral judgment, for Hume, is fundamentally emotional. When we experience a sentiment of approval or disapproval towards something, we become aware of a particular moral duty to either endorse or condemn that thing. Moreover, this sentiment intrinsically motivates us to act in accord with the duty. Reason, Hume explains, plays only an instrumental role in moral judgment, indicating what the best means are for carrying out our emotionally determined ends. Reason, as he puts it, “is, and ought to be the slave of the passions, and can never pretend to any other office than to serve and obey them” (Hume 1967, p. 415).

Smith likewise considers moral judgment to be essentially emotional. His account of moral judgment, however, focuses more on sympathy. For Smith, moral judgments are grounded in our ability to feel sympathetic towards others. Thus to judge that “X” is

wrong is to judge that if we were in the agent's position, we would judge "X" wrong. A moral judgment is appropriate insofar as the emotion anticipated by the individual making the judgment is analogous to the emotional felt by the agent. In Smith's words (1971, p. 22):

When the original passions of the person principally concerned are in perfect concord with the sympathetic emotions of the spectator, they necessarily appear to this last just and proper, and suitable to their objects; and, on the contrary, when, upon bringing the case home to himself, he finds that they do not coincide with what he feels, they necessarily appear to him unjust and improper, and unsuitable to the causes which excite them. To approve of the passions of another, therefore, as suitable to their objects, is the same thing as to observe that we entirely sympathize with them; and not to approve of them as such, is the same thing as to observe that we do not entirely sympathize with them.

Hume's and Smith's sentimentalist views were sharply criticized by Kant. According to Kant's rationalist theory, emotion has no role in moral judgment. Rather, we become aware of moral values by applying the "categorical imperative." While Kant specifies a number of different formulations of the imperative, the most common formulation involves the instruction to "act only according to that maxim whereby you can at the same time will that it should become a moral law" (1993, p. 421). For any given action, then, if you affirmatively judge that its universal acceptance ought to be endorsed, that action is morally good. Simply recognizing the objective value of the action will provide intrinsic motivation to see the act through, so long as the individual is conducting himself rationally.

As a reaction to Kant's thoroughly rational deontological theory, Bentham and Mill revived Epicureanism in their doctrine of utility. Utilitarianism, explains Mill, "holds that actions are right in proportion as they tend to promote happiness; wrong as they tend to produce the reverse of happiness. By happiness is intended pleasure and the

absence of pain; by unhappiness, pain and the privation of pleasure” (Mill 2001, p. 7). The utilitarians believe, contrary to Kant, that the means taken to achieve some end are morally irrelevant (Mill 2001). As long as the consequences of an action, “X”, result in a greater proportion of pleasure than pain, then “X” ought to be judged morally admirable, irrespective of the motive involved. Hence for utilitarians, moral judgment is sentimental insofar as we rely on our emotions to reveal those things that are pleasurable and those things that inflict pain.

1.2 Moral Judgment in Psychology

Psychological accounts of moral judgment have a considerably shorter history than their philosophical counterparts. This is largely due to the fact that philosophy and psychology were considered one and the same discipline until the late 19th century. The disagreement between those who consider moral judgment to be fundamentally cognitive and those who see moral judgment as essentially emotional carried over into the field of psychology.

One of the first psychologists to write extensively about moral judgment is Piaget. His belief is that moral judgment is essentially cognitive. Piaget (1948) argues that there is a parallel between moral and intellectual development. We learn moral norms in the same way that we learn logical norms, through reason. The acquisition of moral norms takes place in two stages: constraint and cooperation. First, as children, we are faced with moral constraint from adults in positions of authority, such as parents or teachers. This constraint leads to heteronomy and consequently to moral realism. Second, after moral rules become interiorised and generalized, we move to a cooperative stage. Cooperation leads to autonomy. At this stage we formulate moral judgments and act purely out of

respect for the moral rules themselves, instead of out of respect for a person who wields authority.

Inspired by Piaget's study of moral development, Kohlberg (1971) devised a more substantial account of how moral judgment becomes increasingly sophisticated as children mature into adults. Kohlberg argues that there are six distinct levels of moral development and that these fall under the scope of preconventional, conventional, or postconventional understandings of morality. At the preconventional level, children respond to cultural labels of right and wrong, but they interpret these values in terms of the consequences that will arise if they are carried out. During this stage, children are primarily concerned with avoiding punishment or receiving praise for their actions and their moral judgments will reflect these aims. At the conventional level, individuals are more concerned with fulfilling the expectations of their community. These expectations are perceived as valuable in their own right, regardless of any immediate or obvious consequences. Not only does the individual aim to conform to personal and societal expectations, but he also will actively maintain, support, and justify them, and will identify with the group or individual responsible for generating the expectations. Finally, at the postconventional level, the individual is concerned with defining moral values that are valid and applicable in an autonomous sense. That is, the individual conceives of moral values holding irrespective of the groups or individuals who hold the values in esteem, and independent of the individual's self-identification with these groups.

Kohlberg's theory of moral judgment is purely rational in the following sense. At the preconventional, conventional, and postconventional stages of morality, moral

judgments are the result of pure, cognitive reflection. In Kohlberg's words (1971, p. 230-31),

...the moral force in personality is cognitive. Affective forces are involved in moral decisions, but affect is neither moral nor immoral. When the affective arousal is channelled into moral directions, it is moral; when it is not so channelled, it is not. The moral channelling mechanisms themselves are cognitive.

Kohlberg additionally notes that the idea that moral judgments are difficult as a result of the conflict between cognition and emotion is misleading. On his account, individuals fail to uphold moral values not because their ability to restrain their emotional impulses is weak, but rather because their cognitive definition of right and wrong is less independent of what other people think (Kohlberg 1971). That is, they have yet to reach the most advanced, postconventional stage of moral judgment where moral values are defended and upheld in their own right.

Though Kohlberg is a rationalist about moral judgment, his former colleague, Carol Gilligan, argues that moral judgment, particularly in women, is typically emotional. Gilligan develops this argument after noticing that women rarely progress past the conventional stage of morality on Kohlberg's scale. The reason that women fail to reach the post-conventional level, explains Gilligan, is because Kohlberg has mistakenly conflated the moral development of men and women (1982). Kohlberg fails to realize that the ethical thought of men and women develop in very distinct ways. Gilligan agrees that a man's moral development is typically grounded in an ideal of impartial justice. A woman's moral development, however, is grounded in responsibility and caring relationships. She explains that women construe the moral problem "...as a problem of care and responsibility in relationships rather than as one of rights and rules". This "ties the development of their moral thinking to changes in their understanding of

responsibility and relationships, just as the conception of morality as justice ties development to the logic of equality and reciprocity” (Gilligan 1982, p.72). Women are thus more likely to base their moral judgments on their feelings, whereas men tend to base their moral judgments on abstract moral norms.

While Gilligan’s sentimentalism applies strictly to women, other writers in the “ethics of care” tradition extend their theories to men as well. For instance, Nel Noddings and Peta Bowden argue that both women’s and men’s moral judgments are a product of our care and concern for the well-being of others (Noddings 1984, Bowden 1997).

1.3 The Contemporary Sentimentalist Debate

Sentimentalist accounts of moral judgment are popular in both contemporary philosophy and moral psychology. In philosophy, this is because sentimentalists have an easier time accounting for moral motivation than rationalists do. For the sentimentalist, emotions have the intrinsic power to motivate people to act in accord with their moral judgments. A judgment is thus causally connected to the corresponding action. Moreover, sentimentalism is consistent with the view that moral values do not have an independent existence, but rather, in some way, depend on us for their value (D’Arms and Jacobson 2000). In moral psychology, sentimentalist accounts of judgment are strengthened by neural evidence showing that emotional brain areas are activated when individuals make moral judgments (Greene and Haidt 2002).

In what remains of this section, I would like to explore a series of contemporary sentimentalist views of moral judgment. These views are sometimes called “neosentimentalist” in the metaethical literature (D’Arms and Jacobson 2000, Prinz 2006). They are sentimentalist insofar as they consider our moral judgments fundamentally

emotional, yet they deviate from simple sentimental theories, since they demand that our sentiments be appropriate or warranted in some way. I think it is important to investigate the ongoing debate between two factions of neosentimentalist writers, the cognitivists and the non-cognitivists, since Prinz formulates his sentimental theory with this particular dispute in mind.

On Prinz's understanding, simple sentimental theories are those that take the judgment "X is wrong" as simply having a sentiment of disapprobation towards X. Although Prinz uses Hume's position as an example of simple sentimentalism, I believe that Hume's view is more complex than Prinz accounts for. It should thus be noted that when I use the term "simple sentimentalism," I am not referring to Hume's position, but rather to any sentimental theory which equates moral judgment with the experience of a sentiment of approval or disapproval.

Simple sentimental theories are traditionally criticized for two reasons (D'Arms and Jacobson, 2000). First, they are incapable of accounting for moral disagreement. That is, they cannot explain why we spend time passionately debating morality if the judgment "X is wrong" simply amounts to "I disapprove of X." Our judgments should be equally correct if all they amount to is a description of our mental states. This problem is taken care of easily enough by expressivism or emotivism, which take the judgment "X is wrong" to be an *expression* of the mental state of disapproval (see, for instance, Ayer 1956). Moral disagreement thus becomes a disagreement over moral attitudes. Debates about our moral judgments can be understood as attempts to influence the attitudes of others (D'Arms and Jacobson 2000).

The second problem, encountered by sentimentalist and expressivist theories alike, is that if negative evaluations of something are taken as expressions of our attitudes of disapproval, then the theory will have no way of distinguishing between negative concepts (e.g., wrong, harmful, ugly, shameful, disgusting, etc.) even though these concepts are importantly different. This problem is dealt with easily enough as long as the sentimentalist distinguishes between particular kinds of disapproval each of which will have a certain emotion that corresponds to it. If this answer is accepted, however, both the sentimentalist and the expressivist are forced to address a much more challenging problem. The new difficulty is that we do not necessarily need to feel a sentiment in order to make the judgment associated with it. Moreover, we can reject judgments, even while experiencing the corresponding sentiment. It is thus neither necessary nor sufficient that we be in the relevant emotional state required to make the related judgment (D'Arms and Jacobson 2000).

The task taken up by the neosentimentalist, then, is to develop a theory that will preserve the notion that our judgments are ultimately grounded in the emotions, but at the same time, explain why our judgments appear to have non-emotional, normative force. Neosentimentalists have attacked this challenge from two opposing directions. Some argue from a cognitivist perspective, while others support a noncognitivist solution. It should be noted that I am using the term “cognitivism” to denote those views that take moral judgments to express beliefs. This should not be confused with the psychological use of the term “cognitive,” which refers broadly to cognitive states such as beliefs, memories, concepts, and so forth. Cognitivists can be either rationalists or sentimentalists about moral judgment. The rational cognitivist holds that moral judgments express beliefs

about objective moral norms, while the sentimental cognitivist argues that moral judgments express beliefs about our sentiments. The particular type of cognitivism I will be discussing is sentimental cognitivism.

According to the cognitivist, moral judgments express beliefs about our sentiments (i.e., of approval or disapproval towards something). Beliefs are to be taken at face value to assert propositions. Our beliefs are thus truth-apt; that is, they are capable of being assessed in terms of truth and falsity. It follows that our moral judgments themselves can be either true or false. Two well-known proponents of this view are David Wiggins and John McDowell.

As Wiggins and McDowell see it, both sentiments and beliefs are required in order to make moral judgments (Wiggins 1987, McDowell 1994, 1998). Whenever we express a sentiment of moral approval or disapproval towards something, the expression of the sentiment will always entail a belief indicating that the thing is indeed morally wrong. Moral judgment thus consists of a sentiment along with a belief affirming that the sentiment merits our approval or disapproval. The belief that the sentiment is warranted is truth-apt; that is, the belief is either true or false. It is this quality that gives our moral judgments an objective gloss. It should be noted, however, that Wiggins and McDowell do not take cognitive judgments such as “wrongness” to be objective or to exist in a mind-independent sense. Rather, such judgments depend upon our perception of them for their value. Their value is something that we project onto them. Our moral judgments are thus still essentially emotional. The explanation is simply not reductive in the same way that traditional sentimental theories are (Darwall, Gibbard, and Railton 1997).

Non-cognitivism, alternatively, denies the need to appeal to any cognitive states such as beliefs in an account of moral judgment. The non-cognitivist, like his expressivist predecessors, maintains that our moral judgments express our sentiments of approval or disapproval towards things. Since these sentiments are not truth-apt, our moral judgments are incapable of being evaluated in terms of truth and falsity. Hence, there are, strictly speaking, no such things as moral facts. If, however, our moral judgments do not express moral facts, then the non-cognitivist must provide an alternative explanation as to why our moral judgments tend to have the same features as claims to objective truth. The non-cognitivist responds by arguing that our moral judgments are “quasi-objective,” that is, there are ways in which our moral judgments mimic objective, factual judgments (Darwall, Gibbard, and Railton 1997). Simon Blackburn and Allan Gibbard formulate distinct explanations as to why our moral judgments appear and are discussed as though they are objective facts.

Blackburn’s (1994, 1998) quasi-realist theory holds that moral judgments express our attitudes of approval or disapproval towards things. Initially, this is done not with moral predicates such as ‘right’ or ‘wrong,’ but rather with emotive predicates like ‘boo!’ or ‘hurray!’ So, when we make a judgment such as “boo! killing,” we are really projecting an attitude of disapproval towards killing onto an objective event in the world. Because these expressions of our attitudes are dependent on facts in the world, they take on an air of objectivity. On the surface they appear to us as though our moral judgments and assertions really are warranted, factual claims, the truth or falsity of which are capable of debate, discussion, etc. While we come to talk about moral claims *as if* they are objective, they are in reality, fundamentally subjective expressions of our sentiments.

Gibbard's (1990) theory of norm expressivism holds that moral judgments express our acceptance of norms. Moral judgments, on Gibbard's account, are judgments about the rationality of our sentiments. An action can be judged morally wrong if and only if it is rational for the agent to feel guilty about having carried out the act, and for others to feel angry or resentful towards the agent for having acted in such a way (Miller 2003). When described in this way, it is not immediately clear how Gibbard's view is a non-cognitivist one. The non-cognitivist aspect of his theory, however, lies in his analysis of rationality. For Gibbard, rationality does not consist in property ascription or truth-aptness. Rather, to say that something is rational is to express acceptance of a system of norms that permit that thing. By "norm," Gibbard's does not mean an objective, mind-independent value. Rather, he explains, a norm is "a possible rule or prescription expressible by an imperative... The main thing to be explained is not what a norm is, but what "accepting a norm" is—or, more precisely, what it is for something to be permitted or required by the norms a person "accepts." I mean these latter notions to be psychological: they are meant to figure in an explanatory theory of human experience and action" (Gibbard 1990, p. 46). Since moral judgments are analysed as rational judgments, and since Gibbard is a non-cognitivist about rationality, moral judgments themselves are non-cognitive.

While it is clear from this brief survey that cognitivists and non-cognitivists clash over the fundamental nature of moral discourse, these neosentimentalist theories remain consistent in one important way. Wiggins, McDowell, Blackburn, and Gibbard all adhere to some version of what D'Arms and Jacobson refer to as the "response dependency thesis" or RDT. The crucial idea behind the RDT is that "to think that X has some

evaluative property ϕ is to think it appropriate to feel F in response to X” (D’arms and Jacobson 2000, p. 729). In other words, the moral judgment that something is wrong amounts to the judgment that the expression of disapproval towards that thing or the belief that the thing is wrong is merited, warranted, endorsed, justified, appropriate, and so forth.

Adhering to a RDT raises a serious problem for all of these neosentimentalist theories. The problem, as D’arms and Jacobson explain it, is that none of these philosophers “has yet sufficiently distinguished the particular species of appropriateness of response that is relevant to property ascription—to whether some X is ϕ ” (2000, p. 731). Put more simply, the neosentimentalists do not explain what sorts of reasons make our moral judgments appropriate. The following example borrowed from D’arms and Jacobson helps to clarify the problem. They write (2000, p. 731):

For instance, imagine that you have a rich and generous but touchy friend, who is extremely sensitive about his friends’ attitudes towards his wealth. If he suspects you of envying his possessions, he will curtail his largesse. That is a good reason not to envy him, if there is any chance whatsoever that you will betray your attitude in action or nondeliberate behaviour, but surely it does not speak to whether his possessions are enviable. Another reason you might think it inappropriate to envy him would be based on moral qualms about being pained at a friend’s good fortune, but this seems to irrelevant to the ϕ property. While such good strategic and moral reasons can count in favour of (or against) feeling some sentiment, they seem like the wrong kind of endorsement or criticism of it. The trouble is that to call a response ‘appropriate’ is vague praise. This is not merely a quibble about how well certain philosophers have chosen or defined their terms. Whatever one’s preferred normative locution, the point remains that only certain good reasons for or against having a response bear on the associated evaluative judgment. Until these reasons are identified, there is nothing to stop sentimentalism from yielding systematically wrong answers to evaluative questions.

This example and the subsequent narrative make it clear how neosentimentalism falls victim to the “conflation problem”; neosentimentalists are faced with the difficult

challenge of explaining how to differentiate (and hence to avoid conflating) moral and prudential reasons for feeling a sentiment from reasons bearing on whether X is ϕ .

Jesse Prinz, in his recent (2006) article titled “The Emotional Basis of Moral Judgment,” takes on the task of developing a sentimentalist account of judgment that is able to avoid the conflation problem. He argues that recent evidence from cognitive neuroscience can be used to confirm the simple sentimentalist claim that moral judgment consists in having a sentiment of approbation or disapprobation towards something. If empirical evidence can back up the simple sentimentalist thesis, then no metacognitive ascent to neosentimentalism is necessary. Since the conflation problem is something that affects neosentimentalist theories but not simple sentimentalist views, Prinz can avoid it altogether.

Although I think that Prinz’s account deserves criticism on metaethical grounds, I will not concern myself with constructing a metaethical critique in this paper. What I will do, however, is investigate the empirical evidence that Prinz presents in support of his sentimentalist theory. My decision to focus on Prinz’s sentimentalist theory is deliberate for the following reason. Prinz’s view exemplifies a trend in contemporary metaethics that supposes that neural evidence linking the emotions to moral decision-making automatically supports a sentimentalist account of moral judgment (e.g., see also Nichols 2004 and forthcoming). I think, however, on closer examination, it becomes clear that the neural evidence has been misinterpreted by Prinz and his contemporary sentimentalist allies.

In the following chapter I will argue that although the empirical data Prinz cites is incapable of supporting his sentimentalist view, his argument nonetheless reveals some

important clues as to how empirical data can effectively be integrated into future discussions about moral judgment. Before proceeding with my critique of Prinz's position, however, it is important to provide a brief exposition of his sentimentalist theory.

Chapter 2: Prinz's Sentimentalist View of Moral Judgment

Prinz's central claim in "The Emotional Basis of Moral Judgments" is that empirical evidence gathered from cognitive neuroscience can support a simple sentimentalist view of moral judgment. Prinz argues additionally that moral facts are response-dependent and that a form of motivational internalism is true. It should be noted that Prinz's use of the term "response-dependent" does not correspond to D'Arms and Jacobson's use of "response-dependent." For Prinz, "response-dependent" is taken to mean that "the bad just is that which causes disapprobation in a community of moralizers" (2006, p. 29). On D'Arms and Jacobson's use of the term, the bad is that which causes disapprobation *appropriately* in a community of moralizers. While the claims that moral facts are response-dependent and that motivational internalism is true are both relevant and important to Prinz's overall view, the critique I offer in the following section will focus on his argument for sentimentalism. Before moving forward to my critique, however, I would like to open this chapter with a brief summary of Prinz's article, so his argument for sentimentalism may be understood in the correct context.

2.1 A Brief Overview of Prinz's Position

In the past, philosophers have attempted to establish a link between emotion and moral judgment through reasoned argument and reflection alone. Prinz, however, argues that we should appeal to empirical data in order to back up our philosophical intuitions about this purported connection. Current evidence, he believes, supports the idea that our moral judgments are fundamentally emotional. Prinz opens his article with a review of some recent findings from empirical psychology and cognitive neuroscience that corroborate three increasingly strong claims about the relationship between emotion and moral

judgment. First, he argues that emotions co-occur with moral judgment. Second, he asserts that emotions influence and are sufficient for moral judgment. Finally, Prinz attempts to show that emotions are necessary for moral judgment. Though Prinz does not specify what exactly he means by “necessary,” Jones (2006, p. 46) interprets Prinz as arguing that emotion is *conceptually* necessary for moral judgment (i.e., necessary in all possible worlds). I disagree with Jones’ rather strong interpretation, however, and take Prinz to be arguing for *causal* or *psychological* necessity instead.

Although Prinz presents a sizable amount of empirical evidence in support of his hypothesis that emotions are correlated to, sufficient for, and necessary for moral judgment, he is aware that empirical evidence alone is not demonstrative of a theory of moral judgment. He believes, however, that he can back up his sentimentalist thesis through further explanation and systematization of these empirical findings. Prinz thus proceeds to present a variation of the Humean sentimentalist view of moral judgment, which, he argues, can be defended with the empirical evidence he has already drawn to our attention. In short, Prinz’s theory says (2006, p. 31):

To believe that something is morally wrong is to have a sentiment of disapprobation towards it.

By “sentiment,” Prinz simply means any disposition to have emotions. To have a sentiment of disapprobation towards something is just to be disposed to have a negative emotional response towards that thing, that is, to disapprove of it.

On Prinz’s account, different types of moral transgressions will elicit different types of emotional reactions. This claim, he argues, can be backed up by anthropological evidence. For instance, Shweder et al. (1997) report that there are three broad categories of moral rules: those designed to protect persons, those designed to protect the

community, and those pertaining to the perceived natural order. These rules, Prinz notes, correspond to different emotions. Crimes against persons incite anger, crimes against the community incite contempt, and crimes against nature incite disgust. Prinz argues moreover, that we experience emotional reactions to varying degrees. The transgressions to which we have a more severe emotional reaction will indicate the corresponding severity of the transgression (Prinz 2006, p. 34).

Similar to McDowell and Wiggins, Prinz argues that our sentiments refer to response-dependent properties. That is, when we feel a sentiment of disapprobation towards something, we ascribe to that thing the property of causing the disapprobation in us. For example, if I say that killing is wrong, I ascribe to the act of killing the property of causing anger, remorse, and so forth, in me. Since our moral judgments refer to response-dependent properties, they are also truth-apt. Prinz's cognitive stance on the nature of moral judgments allows him to avoid the problem facing Blackburn and Gibbard of accounting for the apparent objectivity of moral discourse.

However, since Prinz is defending a simple sentimentalist thesis rather than a neosentimentalist one, his account can still be challenged by the same charges that are laid against simple sentimentalist theories. The challenge, recall, is for the sentimentalist to distinguish between importantly different negative responses (Prinz calls this the problem of error) and to account for moral disagreement. As we saw earlier, the neosentimentalist can easily deal with these problems by introducing the notion of warrant or appropriateness. This metacognitive move, however, leads the neosentimentalist directly into the conflation problem. This is something Prinz wants to

avoid altogether. He thus responds to the problem of error and the problem of moral disagreement without making any metacognitive moves.

First, Prinz takes on the problem of error. The problem, as Prinz sees it, is that if saying something is wrong simply refers to the emotion causing the disapprobation in me then it becomes impossible to judge something wrong in error. This consequence, he notes, can be avoided if we idealize. That is, we can say that wrong refers only to those things that elicit disapprobation in me where I am aware of all of the relevant facts and am free from any emotional biases that are unrelated to the matter at hand. Prinz, however, chooses to endorse a more straightforward reply. His solution depends on drawing a distinction between emotions, which are occurrent states, and sentiments, which are dispositions to experience emotions. Basic moral values, argues Prinz, might consist in linking a sentiment to a specific kind of action in our long-term memory. For instance, we may have a negative sentiment towards cheating. An action, X, may cause an emotion of blame because it has been mistaken for an instance of cheating, although in reality, it is not. This is an error because X is not an instance of cheating. Our judgment that X is wrong is erroneous since we have not established a sentimental policy in our long-term memory linking X with the disposition to experience blame (Prinz 2006, p. 35). According to Prinz, this response to the problem of error allows him to avoid making any metacognitive moves, such as idealizing.

Next, Prinz moves forward to discuss the problem of disagreement. The problem is that if saying something is wrong just means that it is wrong for me, discussions about what is actually wrong will be spurious. Prinz offers several responses to this challenge. First, if we have the same moral values, then what is wrong for you and what

is wrong for me will likely be extensionally equivalent. The debate may turn on a factual disagreement rather than a moral one. For example, a disagreement over abortion may be a disagreement over what it is to be a person, rather than whether abortion itself is morally right or wrong. Second, we have reason to debate moral issues simply because of the practical consequences. That is, we need to come to some consensus about issues like abortion in order to determine how to deal with the related practical issues such as legal rights, and so forth. Debate is a good means for reaching a consensus. Third, we might have some shared basic values which we can use to find some common ground. Fourth, debates about morality can be considered legitimate because of our tendency to project sentimental properties onto the world without realizing or acknowledging that they are response-dependent. Finally, the claim that debates about morality are spurious provides a reason for why moral debates are often interminable; that is, people reside in unique moral spheres (Prinz 2006, p. 36).

Although this marks the end of Prinz's defence of sentimentalism, he argues further that his sentimentalist view receives additional support from the fact that it is capable of explaining three other things that are essential to any account of moral judgment. First, Prinz argues that his theory can explain the link between emotion and motivation. A sentimentalist view can explain why the move from a moral judgment to an action is immediate. In his words (Prinz 2006, p. 36):

If sentimentalism is true, thinking that an action is wrong disposes one to having negative emotions towards it, and negative emotions are inhibitory: they promote avoidance, ceasing, intervention, withdrawal, and, when anticipated, preventative measures. Beliefs about obligations are not add-ons to beliefs about wrongness; beliefs about wrongness carry the motivational force that we experience as being under an obligation. Consequently, moral judgments vie for control of the will. When they occur, we are thereby motivated to act.

Second, Prinz's sentimentalism can explain how we distinguish between moral rules and simple conventions. Moral rules, argues Prinz, are grounded in emotional responses, while conventions generally are not. This is how they are distinguished from one another. Conventions, says Prinz, that are accompanied by some sort of emotional reaction (e.g., things that elicit feelings of disgust) often become moralized. For instance, since we have been emotionally conditioned to respect certain rules of etiquette (e.g., it is wrong to spit in public), violations of these rules tend to be treated moralistically rather than conventionally.

Third, Prinz believes his theory can reconcile intuitionism with sentimentalism. In particular, his sentimentalism can be used to support the intuitionist claim that moral judgments are self-justifying. That is, moral judgments do not require independent argumentative support. Intuitionists typically base their self-justification claim on the phenomenological intuition that moral judgments seem self-evident. Sentimentalism, argues Prinz, can bolster this claim. Sentimental judgments, he explains, generally seem self-evident. For instance, it is evident that gin and tonic is likeable because it induces pleasure when I drink it. Such judgments do not seem to require any more justification than this. Moral judgments, argues Prinz, are no different. They have "a perception-like immediacy that does not seem to require further support" (Prinz 2006, p. 37). Thus, if moral judgments are sentimental and refer to response-dependent properties, then a judgment that something is wrong seems self-justifying because such an act elicits a negative sentiment expressed by the judgment and having the power to bring out such negative sentiments is constitutive of being wrong.

After describing the explanatory fruits of his sentimentalist account, Prinz evaluates whether or not it is possible to make external moral judgments; that is, moral judgments that are not intrinsically motivating. This topic, of course, draws Prinz into the long-standing debate between motivational internalists, who believe that it is a matter of conceptual necessity that our moral judgments are connected with our actions, and motivational externalists, who argue that no necessary connection exists.

Prinz enters the debate in support of the internalist intuition that there is a necessary link between moral judgment and action. He does admit, however, that there are instances where we make dispassionate moral judgments. Prinz argues that his sentimentalism can allow for two sorts of dispassionate moral judgments. The first kind of dispassionate judgment occurs when I have a sentiment but do not manifest it. For example, I love chocolate, and although this sentiment is truly self-ascribed, I do not always experience a rush of pleasure when I bite into a Crispy Crunch bar. Moreover, I can say that killing is wrong even when I am not experiencing feelings of anger or outrage. I can make this type of dispassionate judgment because sentiments are emotional dispositions, they are not, in and of themselves, occurrent in the way that emotions are. It is thus possible to have a sentiment without it being manifested as an occurrent emotion. Though the judgment is not immediately motivating, it expresses a dispositional state that, under the right circumstances, is motivational.

It is also possible to ascribe moral judgments dispassionately. We tend to do this when we are discussing the moral attitudes of others. Consider, for instance, a tribe that endorses polygamy. I could observe the tribe's marriage practices and infer that the tribe members find the practice of polygamy morally commendable. Given the morality of the

group, the tribe members ought to be polygamous. This judgment does not, however, serve to express my own sentiments. I am expressing a dispassionate judgment about the group's morals, not a dispassionate moral judgment (Prinz 2006, p. 38).

Prinz's sentimentalism therefore shows that there is no such thing as a dispassionate moral judgment. All moral judgments are emotionally charged or "hot," to put it in Prinz's words. As a result, moral judgments are either inherently motivating or parasitic (as in the chocolate case) on those that are inherently motivating. This conclusion supports motivational internalism.

Prinz's argument for motivational internalism concludes his account of moral judgment. In this paper he has argued for the following three interrelated claims: sentimentalism is true, moral facts are response-dependent, and motivational internalism is true. Although Prinz makes some questionable metaethical moves when defending the latter two of these claims, I am more concerned with the support he offers in favour of his initial claim, that sentimentalism is true. Unless Prinz can adequately defend simple sentimentalism, his arguments for the response-dependence of moral facts and for motivational internalism cannot even get off the ground. I do not think that Prinz has adequately defended simple sentimentalism. The following section of my paper will be devoted to taking up the reasons why I think this is the case.

2.2 Critique of Prinz's Sentimentalism

The strength of Prinz's sentimentalist theory is in his ability to establish a strong argument for a link between emotion and moral judgment. The evidence Prinz cites in favour of such a link has been taken primarily from experiments in cognitive neuroscience designed to monitor brain activity during moral decision-making and other

morally relevant tasks. Prinz also presents some anthropological evidence, which has been gathered simply through empirical observation of human behaviour. My task in this section is to argue that the empirical evidence cited by Prinz establishes neither a necessary nor a sufficient connection between emotion and moral judgment. As I mentioned earlier, it is my understanding that Prinz is arguing for causal or psychological necessity, rather than conceptual necessity. Establishing a necessary and sufficient connection between emotion and moral judgment is, on Prinz's account, imperative for his claim that moral judgment is fundamentally emotional. I will begin this section by examining Prinz's arguments for the co-occurrence, sufficiency, and necessity of emotion in more detail. Throughout my exposition of these arguments, I will point out some critical errors with his evidence as well as with his arguments in general. Following my critique, I will explain why I agree with Prinz's weaker claims, that emotions often co-occur with and influence moral judgment. The significance of the emotion-judgment connection will be further elaborated in the following chapter.

The claim that emotions typically co-occur with moral judgments is, as Prinz explains, not terribly controversial. Brain scans, such as fMRI or PET scans, confirm that we typically elicit a negative emotional response when we observe a moral norm being violated or when we are required to make moral appraisals. Prinz cites a number of findings as evidence for this phenomenon. For example, Moll, de Oliveira-Souza, and Eslinger (2003) have found that when subjects make moral judgments, brain areas associated with emotional responses are activated. When subjects make strictly factual judgments, however, these brain areas are not activated. In a separate study, Sanfey et al. (2003) monitored brain activity while subjects played an ultimatum game. When subjects

judged that they were being treated inequitably, brain activity was recorded in areas associated with emotion. This indicates that inequitable or unfair treatment elicits a negative emotional reaction. The findings of Berthoz et al. (2002) indicate that brain areas related to emotion are engaged when subjects consider violations of social rules. For example, when told a story about a dinner guest who, after tasting the host's food, unapologetically spat it out into his napkin, subjects showed activation in emotional brain areas. Greene et al. (2001) have discovered that emotional brain areas are activated when subjects are asked to consider moral dilemmas. Finally, Kaplan, Freedman, and Iacoboni (forthcoming) have found that emotional brain areas are engaged when subjects view pictures of politicians whom they oppose (Prinz 2006, p. 30).

None of these findings, Prinz recognizes, are all that surprising. Brain scans lend empirical support to the observation that emotions typically arise in response to morally significant events. Establishing this, however, does not allow Prinz to make any metaethical progress. What he must do in order to advance his metaethical argument is establish the specific role that emotions play in moral judgment. He thus moves forward to defend his second claim, that empirical evidence supports the hypothesis that emotions influence and are sufficient for moral judgment.

Prinz begins by defending the claim that emotions can influence moral judgment. This can be established, he argues, by showing that the experience of a negative emotion can lead to the construction of a more negative moral appraisal than otherwise would have been made. Schnall, Haidt, and Clore's (forthcoming) findings support this hypothesis. They discovered that when subjects are seated in a disgusting environment (e.g., at a desk with a chewed pencil, a used tissue, and a greasy pizza box) they tend to

rate actions more morally wrong than subjects who are seated in a clean environment. Prinz notes, however, that these findings may only establish that emotions can potentially draw our attention to the morally salient features of a situation. What he would prefer to establish is that emotions are sufficient for moral judgment.

Prinz argues that Wheatley and Haidt's (forthcoming) findings provide evidence for the sufficiency claim. In their study, Wheatley and Haidt hypnotized subjects to experience feelings of disgust whenever they heard the emotionally neutral word "often." This word was then integrated into two separate scenarios. The first scenario described a morally reprehensible character, while the second described a morally admirable character. Wheatley and Haidt found that the hypnotized subjects judged the actions of the morally admirable character to be morally wrong when the word "often" was present in the scenario. This, suggests Prinz, shows that the experience of disgust is sufficient to make moral judgments.

Wheatley and Haidt, however, draw a weaker conclusion from these results. In their account of the experiment they write, "...participants used their feelings of disgust...as *information* about the wrongness of an act. This finding indicates that gut feelings can indeed *influence* moral judgments" (forthcoming, p.3, emphasis mine). I agree that this is the most accurate interpretation of their findings. One reason for this is that although the subjects' judgments were prompted by their feelings of disgust, it is doubtful that, were they in a fully informed state, they would maintain their judgment that the morally admirable character was, in fact, morally wrong. Karen Jones also makes this point in her recent critique of Prinz. She writes (2006, p. 48):

Sometimes we think we are responding to a morally relevant feature in the world, but we are actually only looking at the world through the lens of our disgust, or

our shame, and mistakenly supposing the evoking situation to have the kind of morally significant features that would warrant those emotions...Normal subjects could be expected to say that they thought the action wrong or the person reprehensible, but they can now see that there was no basis for this judgment and they were mistaken. Subjects would not take the mere fact of their feeling disgust to warrant the moral judgment.

That subjects will, in all likelihood, retract their initial claim in a state of full information lends support to the idea that emotion can influence moral judgment, though not to the notion that they are sufficient for moral judgment.

It is clear, moreover, that Wheatley and Haidt have a different idea from Prinz of how disgust can influence our moral judgments. They argue that individuals use their feelings of disgust to inform their moral appraisals. This does not entail, however, that this feeling is the only thing informing their judgments. This provides additional reinforcement for the idea that emotions have the power to influence moral judgment. It does not, however, support the claim that emotions are sufficient for moral judgment.

Prinz moves forward to cite Murphy, Haidt, and Björkland's (forthcoming) discovery that subjects judge incest to be morally wrong even when they cannot provide any independent justification for their belief. The intention of the experimenters was, once again, to study the relationship between disgust and negative moral appraisals. In their experiment, Murphy, Haidt, and Björkland asked their subjects to justify their belief that consensual incest between siblings is wrong. For every justification the subjects offered (e.g., because the siblings might have a child with birth defects), Murphy would inform them that their reason was irrelevant to the case (e.g., they used two forms of birth control). The experimenters report that even when the subjects could think of no more reasons justifying their belief, they still insisted that incest was wrong just because it is disgusting. That the subjects could not justify their belief rationally yet continued to

maintain it is, according to Prinz, evidence that their disgust was sufficient for the moral judgment.

I disagree with Prinz for two reasons. First, Prinz ignores the fact that a portion of the subjects *did* give up their belief that consensual incest is wrong when Murphy dismissed the reasons they offered in support of their initial judgment (2006, p. 31). Though this does not disprove Prinz's sufficiency claim, it raises the question of why an emotion would be sufficient for moral judgment in some, but not all cases. This is an issue that Prinz simply does not address.

Second, it is possible for a subject's initial response to a moral transgression to be purely emotional. This does not entail, however, that the subject's response is not guided by reasons. It is more likely the case that the subject assumes that her emotions are drawing her attention to a morally relevant feature of the situation, which, at present, she is incapable of articulating (Jones 2006, p. 50). It is not unusual, moreover, for emotions to play this role. Our affective reactions often draw our attention to important features of situations without explicitly informing us of the reasons for this. Adolphs (2003), for example, argues that disgust is a reaction that evolved in humans to indicate when to avoid things that pose a potential threat to our physical health or the health of our offspring (e.g., contaminated foods, filthy environments, or incestuous relationships). When we experience disgust firsthand, this is, perhaps, the reason why. This reason, though not immediately present to our consciousness, does not mean that it is nonexistent. The reason simply has yet to be articulated. Thus, although the majority of subjects resorted to defending the claim that consensual incest is wrong simply because it is disgusting, it does not follow from this that their judgment is purely grounded in affect

alone. It is entirely possible their disgust is simply drawing their attention to a set of non-articulated reasons. Although this analysis poses a threat to Prinz's sufficiency argument, it is consistent with his weaker claim that emotion influences moral judgment.

There is one final point I would like to bring up regarding Prinz's evidence for sufficiency. Suppose, for a moment, that we grant that these disgust-based studies support Prinz's claim. Disgust is sufficient for moral appraisals. What exactly would this prove about the relationship between emotion and moral judgment? Prinz provides no evidence that the experience of any emotions other than disgust would lead to the same effect. In fact, since emotions of disgust and anger are linked to a completely different brain region than emotions of sadness or fear (i.e., disgust and anger are associated with activity in the insula while sadness and fear are associated with activity in the amygdala), there is, I think, quite good reason to doubt that these findings could be repeated in studies of alternative emotions. Thus, even if disgust is sufficient for moral judgment, it is far from clear that emotion, in general, will be.

I will now move forward to discuss Prinz's claim that emotions are necessary for moral judgment. Prinz provides evidence for this claim in three separate ways. First, Prinz argues that emotions are necessary for moral development. This, he thinks, is because the techniques that parents use to convey moral rules all engage emotions. These techniques (e.g., threatening the child with harm to elicit fear) condition the child to experience negative emotions in conjunction with misdeeds. The emotional response (e.g., fear) serves the purpose of instilling the rule that is to be obeyed in the child.

This piece of evidence, I argue, is irrelevant to necessity. What this example does show is that negative emotions typically accompany the infraction of moral norms when

they are conditioned to do so. This example is not successful, however, in demonstrating that children must experience affective reactions in conjunction with moral infractions, nor does it bolster the claim that emotions are necessary for moral judgment.

Prinz thinks that stronger support for the link between emotion and moral development can be derived from Blair and his colleagues' research on psychopathy (1995, 1997, 2001, 2002). Psychopaths, explains Prinz, have been found to rarely experience negative emotions such as fear or sadness and have trouble recognizing these emotions in the speech and facial expressions of others. According to Blair and his colleagues' experiments, psychopaths experience pain less intensely than normal subjects and are not disturbed by emotionally charged images in the same way as non-psychopathic subjects (1995, 1997, 2001, 2002). This evidence, argues Prinz, indicates that psychopathy results from a low-level deficit in negative emotions. In the absence of negative emotions, psychopaths cannot experience empathetic distress, remorse, or guilt. Since these core emotions are necessary for the normal development of moral concepts, psychopaths are incapable of acquiring the ability to distinguish between conventional and moral misdeeds. According to Prinz, this evidence supports the thesis that emotion is necessary for moral development. This, in turn, lends indirect support to the idea that emotions are necessary for moral judgment in general.

The principle difficulty with Prinz's psychopathy evidence is that he conveniently omits findings which contradict his claim. While Blair did discover that psychopaths find it difficult to recognize facial expressions and sounds related to fear and sadness, he also found that they recognize emotions such as anger and disgust as readily as non-

psychopathic subjects. In a recent study, for instance, Blair writes (2001, p. 493, emphasis mine):

Psychopathic adults and children with psychopathic tendencies show reduced skin conductance to sad, but *not angry expressions*. Moreover, children with psychopathic tendencies have been found to show selective recognition difficulties for sad and fearful expressions but *not for angry, disgusted, surprised, or happy expressions*.

The reason that psychopaths are incapable of recognizing anger and disgust but not fear and sadness is perhaps because these emotions are associated with different areas of the brain. Psychopathy is a disorder related to underdevelopment in the amygdala. Since fear and sadness are associated specifically with amygdala function, it makes sense that these emotions are impaired. Anger and disgust, however, are associated with activation in the insula (Wicker et al. 2003, p. 655). Because the insula is not impaired in psychopaths, a psychopath's recognition of anger and disgust should, likewise, be unimpaired (Blair 2002, p. 682). Prinz, recall, argues that emotions are necessary for moral development. This is meant to support the fact that emotions are necessary for moral judgment. If psychopathic individuals process and experience emotions such as anger and disgust normally then their moral development and moral judgment, in relation to these emotions, ought to be unimpaired. This, however, is not the case. The fact that psychopathic individuals have an underdeveloped capacity for moral decision-making and judgment, even in cases related specifically to emotions like anger, seems to indicate that emotions may not necessarily play a role in moral development or moral judgment. At the very least, this evidence has the power to directly undermine Prinz's sufficiency claim.

Prinz defends his claim that emotions are necessary for moral judgment in a second way by contending that emotions are necessary in a synchronic sense. This means that when we sincerely judge that something is wrong we are necessarily disposed to have a negative emotion towards that thing. For example, although it is possible to say that killing is wrong and to justify this assertion with non-emotional reasons (e.g., it diminishes utility), it is improper to say that someone actually believes killing to be wrong in the absence of a strong negative sentiment towards killing. Alternatively, if someone has a strong sentiment of disapprobation towards killing, it would be proper to say that the individual believes killing is wrong even if they cannot rationally justify their belief. Though Prinz offers no empirical evidence as support for this claim, he is confident that neuroscientific trials will be developed in the near future to test and confirm his thesis that emotions are necessary for moral judgment.

Although Prinz admits that there is no empirical evidence to appeal to in support of this claim, I argue that there is an abundance of empirical evidence supporting the opposite conclusion that we often sincerely judge things to be wrong without experiencing any obvious negative affect. Moreover, we can have a strong feeling of disapproval towards something without believing that it merits a negative moral appraisal. Evidence for both of these claims can be drawn from everyday experience.

In the case of unemotional judgments, consider the activity of voting. Some individuals, of course, will vote for a specific candidate for emotional reasons. Perhaps one of the candidates is a strong supporter of capital punishment while his opponent is firmly against the death penalty. If the idea of capital punishment elicits a feeling of distress or disgust within the voter, they may choose to vote for the candidate who

opposes the death penalty. In this case, the voter is making an overtly emotional judgment. But there are also individuals who decide which candidate to vote for in an unemotional, calculated manner. If the voter has no emotional connection to any particular candidate or their policies then the voter will probably make their decision by weighing one candidate's social or economic plans against the other candidate's proposals in a cold, mathematical manner. This is not to say that the eventual judgment will be entirely unemotional. The example is simply meant to demonstrate that our moral judgments need not be obviously emotional or attributable to explicitly emotional reasons.

Good examples can also be found in favour of the corresponding claim that we can strongly disapprove of something without making a negative moral appraisal. Consider, for instance, smoking. Every time I see someone light up a cigarette, I feel sincerely disgusted. I do not, however, consider smoking to be morally reprehensible. In fact, I do not find smoking morally relevant at all, even though it has the power to cause me to experience a feeling of negative affect. Smoking is, in my opinion, simply gross. These two examples, it seems, contradict Prinz's necessity claim. It is possible to have a sincere emotional reaction to something without making a corresponding moral appraisal. It is likewise possible to form unemotional moral judgments. This supports the conclusion that emotion is not necessary for moral judgment.

Finally, Prinz argues that his necessity claim is strengthened by the fact that emotions are necessary for the establishment of moral values. If moral judgments were based on something like reason or observation, then we ought to see more moral convergence across cultures. Reason and observation, contends Prinz, tend to converge

over time. The fact that moral values are incredibly divergent cross-culturally indicates that moral values do not have a purely cognitive source. The diversity of moral values therefore provides indirect evidence that emotions are necessary for the construction of moral values (Prinz 2006, p. 33). Since a grasp of moral values plays an integral role in moral judgment, it follows that emotions are necessary for moral judgment as well.

Why Prinz believes that reason and observation lead to convergence over time is unclear. First, he does not explain exactly what he means by saying that reason and observation converge over time. Moreover, even if we accept that they do converge, Prinz does not explain how this convergence would be measured. Second, and more importantly, Prinz's presupposition that reason and emotion are distinct is empirically ungrounded. So, too, is his claim that emotion is divergent and reason convergent. Prinz is unsuccessful in establishing any definitive link between emotion and moral diversity.

My examination of the empirical evidence marshalled for Prinz's sufficiency and necessity arguments has hopefully made it clear that his account is fundamentally flawed. My critique has also shown, however, that not all of Prinz's claims are inaccurate. I think that there is good reason to accept his argument that emotions often co-occur with and influence moral judgment. Evidence for this can be found in cognitive neuroscience, where experiments reliably show that tasks related to moral judgment and decision-making activate areas of the brain that are associated with emotional experience (see, for instance, Greene and Haidt 2002, Casebeer and Churchland 2003).

This additional evidence, combined with the evidence Prinz's offers is, I believe, ample to support the claim that emotions often co-occur with and influence moral judgment. Neural evidence does not entail, however, that emotions are either necessary or

sufficient for moral judgment. Since both simple sentimentalist and neosentimentalist theories of judgment demand a causal role for emotion when making moral appraisals, the conclusions that I have drawn in this section appear to pose a problem for sentimentalist views in general. Should these neuroscientific findings then be taken alternatively, to support a rationalist account of moral judgment? In the final section of my paper, I will argue that neural evidence does not favour a rationalist account of judgment any more than it does a sentimentalist one. This, I think, is because neuroscience presents a very different account than traditional philosophical and psychological theories do of what emotions are. By discussing an alternative view of emotions, I hope to show how neuroscience could have a positive effect on future debates about moral judgment in metaethics. Improvement and progress in metaethics, however, will likely require philosophers to relinquish some of their dogmatic beliefs about the nature of reason, emotion, and moral judgment in general.

Chapter 3: Implications for Future Theories of Moral Judgment

In the previous section, I outlined and criticized Prinz's theory of moral judgment. I have argued that moral judgment is often accompanied by emotions and capable of being influenced by emotions. This much seems undisputable. What I have argued against is Prinz's assertion that emotions are necessary for moral judgment and that emotions are sufficient for moral judgment. In my critique, I focused primarily on criticizing Prinz's interpretation and use of neuroscientific evidence. There is, however, an additional reason for criticizing Prinz's necessity and sufficiency arguments. This criticism is related to assumptions he makes regarding the nature of emotion. In order to explain why Prinz's sentimentalist thesis fails in this respect, I must first provide a brief outline of his theory of emotion.

3.1 Prinz's Theory of Emotion

Prinz's view of emotional consciousness is not revolutionary. It is a variation of the view proposed by James and Lange (1922) which has recently been revived by Damasio in what he calls the "somatic-marker" hypothesis (1994). According to this tradition of thought, emotions are simply "perceptions of patterned changes in the body." They are, as Prinz puts it, "gut reactions" (Prinz 2004, p. viii). Views that claim that emotions are fundamentally feelings of changes in our physiology are typically referred to as *somatic* theories of emotion. Alternatively, views that argue that emotions are attributable to some sort of disembodied judgment or propositional attitude are typically called *cognitive appraisal* theories of emotion.

Prinz labels his view an "embodied appraisal theory of emotion." An appraisal, explains Prinz, is "a representation of the relationship between an organism and its

environment that bears on well-being” (2004, p. 51). Cognitive appraisal theories, of course, tend to view this representation as a disembodied mental event. Prinz disagrees. He argues that such representations can be “inextricably bound up with states that are involved in the detection of bodily changes” (Prinz 2004, p. 52). Prinz’s aim is to develop a theory which accepts that emotions are appraisals, yet at the same time asserts that appraisals are grounded in our embodied perception of physiological changes.

In order for Prinz to establish that emotions are appraisals, he must first establish that emotions are mental representations. A mental representation, according to the Dretskean account that Prinz defends, is just a mental state that satisfies two conditions. First, the mental state will carry information about something that reliably causes it to occur. Second, the mental state must be capable of being erroneously applied (Prinz 2004, p. 53). Consider, for example, my concept representing dogs. My concept of dogs is reliably caused or set up by my various encounters with dogs. A mental state reliably caused by dogs, however, could also be reliably caused by a similar animal, like a wolf. Since dogs and wolves look similar, any mental state caused by a dog could also, on occasion, be caused by a wolf. As a consequence, a dog concept will carry information about both dogs and wolves. It is thus possible to fall into error when my dog concept is activated in response to a wolf. Since, however, my dog concept was acquired specifically for the function of carrying information about dogs, it represents dogs and dogs only. If a wolf activates my dog concept, that response is considered erroneous and does not count as a representation.

Prinz’s task is to show that emotions can act as representations in a similar fashion. He must explain how emotions can carry information and that they can likewise

be erroneously applied. Emotions, Prinz argues, are reliably caused by our perception of bodily changes. Emotions are “states within our somatosensory systems that register changes in our bodies” (Prinz 2004, p. 57). Being able to register changes in our bodies, however, is not the same as representing changes in our bodies. So, if emotions do not represent changes in our physiology, what do they represent? According to Prinz, emotions represent things external to us, or more specifically, relations between external states and ourselves. Emotions represent what Prinz calls “organism-environment” relations (2004, p. 60).

This does not mean that emotions represent particular events or things. Rather, argues Prinz, they represent the “formal objects” of emotion. The formal object of an emotion “is the property in virtue of which an event elicits an emotion” (Prinz 2004, p. 62). Consider, for example, my expression of sadness over the death of a loved one. The emotion, sadness, represents the loss of something that I value. The property of loss is the formal object of my emotion. The event that elicits the emotion, the death of a loved one, for instance, is the “particular object” of my emotion. Emotions represent the formal but not particular objects.

On Prinz’s interpretation, it follows that emotions represent “core relational themes.” Core relational themes are “relational properties that pertain to well-being” (Prinz 2004, p. 66). Since emotions represent core relational themes, that means they are reliably caused by core relational themes. Consider, once more, the example of feeling sadness over the death of a loved one. The emotion, sadness, is reliably caused by the loss of something that I value. Loss is a relational property that links my well-being to some thing or event external to me that is of value, e.g., the death of a loved one.

This raises the question, however, of why emotions are set up for this purpose. Why do emotions have the function of being caused by core relational themes? According to Prinz, it makes sense for emotions to have this function from an evolutionary perspective. Emotions track core relational themes because it helps confer a survival advantage and is consistent with individual needs and interests (Prinz 2004, p. 66). Although emotions can be reliably caused by both bodily changes and core relational themes, our emotions only detect core relational themes. Their function is to represent core relational themes, in the same way that a dog concept functions to represent dogs and dogs alone. Emotions do this in spite of the fact that they, like dog concepts, can be reliably caused by other things, such as physiological changes or seeing a wolf. Prinz explains further that (2004, p. 68):

The dog genome, or any other property essential to being a dog, can be called the “real content” of a dog concept. The features by which we detect dogs can be called the “nominal content” of our dog concepts...[Emotions] represent core relational themes, but they do so by perceiving bodily changes. Core relational themes are the real contents of emotions, and bodily changes are their nominal contents.

Prinz distinguishes between the real contents and the nominal contents of emotions in order to reinforce the notion that core relational themes do not carry out a *descriptive* function. In Prinz’s words (2004, p. 65):

Core relational themes do not capture the structure of our emotions or the structure of any other mental representations that are necessary concomitants of emotions. We can form the judgment that there has been an irrevocable loss, but we seldom do. Sadness can occur without that judgment. But sadness represents what the judgment represents. It has the same meaning but a different form.

Emotions thus function to represent core relational themes without explicitly describing them. They track physiological states that reliably co-occur with significant organism-environment relations. It follows that emotions also reliably co-occur with organism-

environment relations. To recap, then, emotions both monitor internal physiological changes and detect external dangers, threats, losses, and other matters of significance. Emotions use our bodies to convey information about how we are faring in the world (Prinz 2004, p. 74).

Prinz's final task is to explain what sort of inner state detects the external object or event and then causes the physiological change to take place. In other words, he must isolate the inner cause of emotions. Prinz argues that the inner cause of emotions is some sort of perceptual state. Even learned emotional responses are generally perceptually mediated. So, for example, if you experience disgust at the sight of a food that made you sick in the past, this is because your brain has associated a perceptual state (e.g., seeing the food) with an emotional response that occurred in conjunction with that perceptual state (e.g., disgust) (Prinz 2004, p. 75). Likewise, emotions can also be triggered by cognitive states like mental images, since cognitive states are, Prinz argues, nothing more than mental representations under organismic control. A link will be formed in memory between an emotion and the representation of the particular object that elicited it (Prinz 2004, p. 75). In summary, Prinz writes (2004, p. 240):

Having an emotion is literally perceiving our relationship to the world...They can deliver information that helps us assess how we are faring. They often allow us to pick up this information before we have made any pertinent judgments...When we listen to our emotions we are not being swayed by meaningless feelings. Nor are we hearing the cold dictates of complex judgments. We are using our bodies to perceive our position in the world.

3.2 Problems with Somatic and Cognitive Theories of Emotion

I have already argued against Prinz's interpretation of the neural evidence he uses to bolster his claim that emotion is both necessary and sufficient for moral judgment. I have

also argued affirmatively that emotions often co-occur with and influence moral judgment. I will now explain in more detail, how and why I think emotions are important for moral judgment. The plausibility of my explanation will depend largely on the account of emotion that underlies my ideas about moral judgment.

In the previous section I offered a brief overview of Prinz's theory of emotion. Prinz's view, like the views of his predecessors, James, Lange, and Griffiths (1997), is a somatic theory of emotion. Prinz, however, labels his view an "embodied appraisal theory of emotion," which is slightly misleading. For Prinz an appraisal is just an organism's perception of the relationship between its self and its environment that is relevant to its well-being (2004, p. 51). This differs drastically from the sense of appraisal intended by those who endorse cognitive appraisal theories of emotion. In the context of these theories an appraisal represents some sort of mental event such as a judgment or a belief (see, for instance, Nussbaum 2001, Oatley 1992, 2004, Lazarus 1994, Scherer 1998, Ortony, Clore, and Collins 1998). Thus, although Prinz alludes to the notion that his theory integrates aspects of both somatic and cognitive appraisal theories, it is nothing more than a basic somatic theory. Emotions, on his account, are just perceptions of changes in our bodily state.

The problem with somatic theories like Prinz's is that they exclude or deemphasize important cognitive aspects of emotional experience. Emotions are produced not merely by perceiving our changing internal physiological states or our changing external relationships with the environment. Our perceptions of our physiological states and of our surroundings are, of course, integral to the production of emotions. Somatic theories, however, fail to explain how the brain makes use of this

information in order to assess how we are faring in the world. Somatic information requires cognitive factors such as attention, judgment, memory, belief, desire, and reason to accumulate, disentangle, and interpret bodily inputs.

Cognitive appraisal theories, unfortunately, fare no better. While these theories do emphasise the important role that thoughts and other cognitive factors play in emotions, they leave no room for physiological factors in their accounts. Though they allow that bodily changes or environmental changes often play a part in emotional experiences as a whole, they argue that the fundamental aspects of emotion (i.e., the ability to assess one's overall situation) are strictly cognitive.

Prinz argues that moral judgment is fundamentally emotional. Emotions, he says, are both necessary and sufficient for moral judgment. This claim, however, must be considered in light of my discussion of Prinz's view of emotion. First, his view that emotions are merely perceptions of changing bodily states is too simplistic to serve as a foundation for a theory of moral judgment. Moral judgment is too complex a phenomenon for a somatic theory of emotion to account for alone. Second, I think that Prinz is mistaken in employing terms such as "necessity" or "sufficiency" in this type of debate in the first place. Using these terms reinforces the philosophical presupposition that the relationship between emotion and moral judgment must be a logical one. It moreover reinforces the philosophical presupposition that emotion and reason are fundamentally distinct.

I think that it is more appropriate to view the relationships between reason, emotion, and moral judgment in terms of degrees as opposed to absolutes. For instance, some moral judgments might be overwhelmingly emotional. This, however, does not

mean that they are not cognitive. Other judgments might be highly calculated, but this does not mean that emotion plays no role whatsoever in the process. By avoiding logical presuppositions such as “necessity” or “sufficiency”, room is left for cases where, perhaps, an individual learns (e.g., after years of meditation) to make a moral judgment that is completely unemotional.

Developing a theory of moral judgment that is capable of integrating cognitive and somatic factors is a possibility. Such a theory, however, will require an account of emotion that integrates aspects from both cognitive appraisal theories and somatic theories of emotion. These theories are not, in fact, inconsistent as many of their proponents believe. With an integrated account of emotion, it is possible to defend a theory of moral judgment that views emotions as vital for making moral appraisals. Before explaining in detail how such a proposal is possible, I would like to explore what an integrated account of emotion might look like.

3.3 The Solution: An Integrated Theory of Emotional Consciousness

A possible candidate for resolving the tension between somatic and cognitive theories of emotion is Thagard’s (forthcoming) neurocomputational account of emotional consciousness. On this account, emotions are “neural processes that represent the overall cognitive and somatic state of the organism” (Thagard forthcoming, p. 1). Thagard’s model, EMOCON for short, explains how different brain areas integrate perceptions of an individual’s physiological states with cognitive appraisals of their current situation through working memory, in order to produce conscious, emotional experience. EMOCON includes the following components: representation, sensory processes, cognitive appraisal, and working memory.

A model of emotional consciousness must include an explanation of how the brain is able to represent the world, bodily states, and its own representations. Thagard suggests that Eliasmith and Anderson's (2003) account of representation is consistent with his theory. As Thagard explains it, the world has a causal effect on an organism's sense organs, which produces neural signals that generate patterns of firing in groups of interconnected neurons (i.e., neural populations). A neural population represents objects or events in the world by means of the causal correlation between its firing patterns and the object or event (Thagard forthcoming, p. 7).

In the same way that organisms have sense organs to detect outer states, they also have internal sensors that detect inner bodily states. Neural populations represent inner bodily states in basically the same way that they represent external aspects of the world (Thagard forthcoming, p. 8). The only difference is that the neural signals are not generated directly by an external cause.

Finally, it is also possible for neural populations to respond to alternate neural populations in order to generate more complex representations. Complex representations are necessary in order for an organism to be capable of the kind of higher-order thought that is required for more sophisticated decision-making and deliberative tasks. It should be noted additionally that these neural populations do not fire signals in one direction only (Thagard forthcoming, p. 9). Rather, the brain is made up of feedback connections, which enable neural populations to send signals in both directions (e.g., signals can be sent from a neural population back to an organism's internal or external sensors).

According to Thagard, a model of emotional consciousness must also include an explanation of the brain areas related to emotions and how they interact. Thagard's

explanation begins with an explanation of Litt, Eliasmith, and Thagard's (2006, forthcoming) model of emotional decision making. The model is known as ANDREA, for short. In their view, all decision making "has an emotional component that involves the interaction of at least seven major brain areas that contribute to valuation of potential actions" (Thagard forthcoming, p. 9-10). These brain areas are the amygdala, the orbitofrontal cortex (OFPFC), the anterior cingulate cortex, the dorsolateral prefrontal cortex (DLPFC), the ventral striatum, the midbrain dopaminergic neurons, and the serotonergic neurons found in the dorsal raphe nucleus of the brainstem. For brevity's sake, I will not elaborate on the specific functions of these emotional brain areas, though their roles are thoroughly explained in a variety of psychological studies (see, for instance, Rolls 2005).

In addition to the seven brain areas modeled in ANDREA, the ventromedial prefrontal cortex (VMPFC) and the hippocampus will also play a role in an account of emotional consciousness. The VMPFC provides important connections between the orbitofrontal cortex and the amygdala. The hippocampus is related to memory and the ability to contextualize in decision making. These two additional areas are represented in Wagar and Thagard's (2004) model known as GAGE. Last, Thagard argues that Morris' (2002) findings related to the thalamus and insula are relevant to a model of emotional consciousness. The thalamus and insula are important because of their role in sending and receiving somatic information (Thagard forthcoming, p. 12).

Any account of emotional consciousness must also give an explanation of how and where cognitive appraisals are performed in the brain. That is, how and where does the brain make use of perceptual and somatic information, as well as accumulated

knowledge, in order to assess its overall current state? Thagard notes that Nerb's (forthcoming) model construes appraisal "as a kind of parallel constraint satisfaction accomplished by artificial neural networks using localist representations" (Thagard forthcoming, p. 12). Emotions and goal-relevant elements are represented in this context by single artificial neurons. Although Thagard agrees that appraisal should be construed as some sort of parallel constraint satisfaction, he does not endorse a localist account of representation. For this reason, Thagard chooses to support Aubie and Thagard's (forthcoming-a) program, NECO, which is able to "show how parallel constraint satisfaction can be performed by distributed representations using large numbers of neurons" (forthcoming, p. 13). A distributed model such as NECO, of course, is more neurologically realistic than a localist representation such as Nerb's. Aubie and Thagard (forthcoming-b) are presently extending their distributed model to perform appraisal of emotional situations, "including both cognitive appraisal with respect to goals and somatic information provided via the amygdala and insula" (Thagard forthcoming, p. 13).

The final component needed for a plausible theory of emotional consciousness is an account of working memory. Thagard describes three candidates for such an explanation. First, Smith and Jonides (1999) propose that working memory "involves both short-term storage of different kinds of information in different brain areas and executive processes of selective attention and task management that involve the anterior cingulate and dorsolateral prefrontal cortex" (Thagard forthcoming, p. 13). Second, Eliasmith and Anderson (2003) "describe how working memory can be modelled by transformations of neural representations" (Thagard forthcoming, p. 13). Third, Aubie

and Thagard (forthcoming-a, b) argue that working memory can be used to “provide a binding between cognitive and affective representations” (Thagard forthcoming, p. 14).

After describing the various mechanisms and components necessary to explain various emotional phenomena, Thagard explains how these elements are combined in EMOCON in order to produce an integrated account of emotional consciousness. His neurocomputational model incorporates ideas from the ANDREA, GAGE, and NECO models as well as Morris’ (2002) observations about sensory inputs. Thagard conjectures that emotional experience results from interactions among all of the elements related to representation, sensory processes, cognitive appraisal, and working memory (forthcoming, p. 14). The interactions between these various elements are represented below in Figure 2.

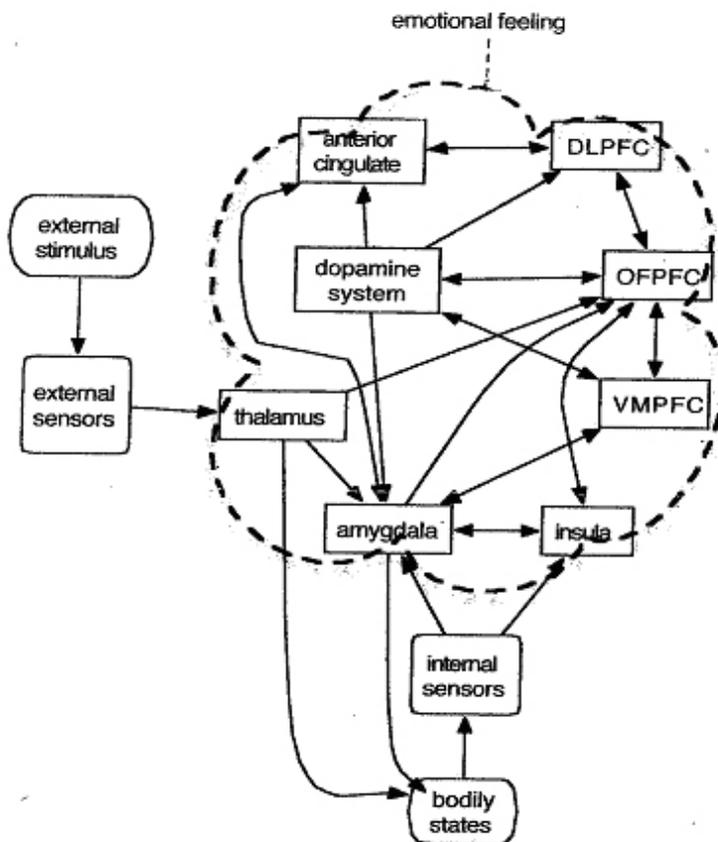


Figure 2: The EMOCON model of emotional consciousness.

Reproduced with permission from Thagard (forthcoming, p. 15).

The EMOCON model illustrates how emotions are not merely perceptions of changing physiological states in relation to our environment. Emotions, moreover, are not merely cognitive appraisals of our current overall situations. Rather, writes Thagard, emotions are patterns of neural activity in the whole system. This includes inputs from internal bodily states and external senses (forthcoming, p. 16). It should also be noted that the diagram presents a number of feedback loops. This is important because it emphasises the fact that emotional consciousness is not represented as an output from any particular brain area. Rather, the diagram is designed to show emotional consciousness as simply “the overall neural process that takes place in interacting brain areas” (Thagard forthcoming, p. 16).

Before moving forward, I would like to clarify that I have chosen Thagard’s theory merely as an example of how an integrative account of emotional consciousness might be possible. A degree of speculation surrounds some of the mechanisms within the EMOCON model. My purpose, however, is not to offer a critical analysis of Thagard’s view, but rather to highlight how an integrative theory might be developed to resolve the difficulties that face cognitive appraisal theories and somatic theories respectively. I will now explain how I think such a theory can have a positive effect on debates about the nature of moral judgment.

3.4 How Might an Integrated Theory of Emotion Impact Future Debates over the Nature of Moral Judgment in Philosophy and Psychology?

I believe that an integrated account of emotional consciousness makes it possible to defend a philosophical or psychological stance that takes emotions to be a vital component of moral judgment. If emotion is understood as Thagard proposes (i.e., simply

as “neural processes that represent the overall cognitive and somatic state of the organism”), it becomes clear why emotion must play an integral role in moral judgment. Emotion is vital for moral judgment because all judgments, moral or otherwise, are typically emotional.

In order to better understand why moral judgment is typically emotional, I think it will be helpful to examine the notion of moral judgment itself. What actually occurs when we are making a moral judgment? Initially, of course, we either perceive an event unfolding in the world or recall an event that previously occurred. After perceiving or recalling such an event, the event becomes represented within the brain’s working memory. The representations of the event begin cycling through feedback loops in the brain, combining with representations about our changing physiological states, representations of beliefs, representations of goals, along with a variety of other sorts of representations. The cycling of these representations between neural populations in the anterior cingulate cortex, dorsolateral prefrontal cortex, dopamine system, orbitofrontal cortex, ventromedial prefrontal cortex, insula, amygdala, and thalamus is the basic stuff of conscious thought, decision-making, and judgment. This cycle is likewise the basic stuff of emotional consciousness. The process that generates moral appraisals is simultaneously generating emotions. Moral judgments just *are* emotional experiences. Emotion is thus vital for moral judgment because when a brain generates a moral judgment it will be generated from the same circuit that also generates emotional experience.

If we accept that emotion is vital for moral judgment, might this imply a victory for sentimentalism? The short answer is probably not. How, then, will an integrative

theory of emotional consciousness affect contemporary debates about the nature of moral judgment? This question can be answered in a number of ways.

First, adopting a view that takes emotion and cognition to be intrinsically linked in consciousness has the power to bridge the longstanding divide between rationalism and sentimentalism in both philosophy and psychology. This is certainly not to say that either side has been defeated. Rather, each side can be confident in the fact that they have gotten part of the moral judgment story correct. Introducing an integrative theory of emotional consciousness to the debate will enable thinkers on either side to understand that they need not adhere exclusively to one view or the other. Accepting that cognition and emotion coexist rather than conflict will open the door for a number of new collaborations within these fields of study. This will allow progress to be made on other underexamined aspects of moral judgment. An integrative account of emotional consciousness will enable us to move past debates over whether moral judgment is fundamentally cognitive or sentimental. It has the power to encourage and inspire new discussions about morality in general.

A second upshot of an integrative account of emotional consciousness is that it has the potential to influence debates related to moral motivation. The literature on moral motivation has traditionally considered emotion to be inherently motivating. This has generally given sentimentalist theories of judgment the upper hand when it comes to explaining action. If, however, we accept that moral judgments are typically emotional, it follows that we must also reconsider our views about moral motivation within this new framework and in absence of any philosophical presuppositions.

Finally, adopting a novel view of emotion may encourage both philosophers and psychologists to develop empirically informed theories of moral judgment that are consistent with contemporary research in both experimental psychology and cognitive neuroscience about the nature of consciousness. It is important, however, that researchers in the humanities strive to integrate empirical data in their theories in an objective, non-biased manner. Experimental results can be a powerful form of evidence for one's theoretical intuitions. Empirical evidence, however, is terribly ineffective when it is misinterpreted or used selectively.

Conclusions

My intention in writing this thesis was to draw attention to a developing trend in ethical philosophy. This trend is to use empirical evidence about the human brain to support philosophical explanations of moral reasoning. The difficulty that arises is that writers, such as Prinz, for instance, have a tendency to misinterpret neural findings or exclude findings from their theories that might contradict their own views. They are providing subjective interpretations of empirical data so that it fits with their ethical presuppositions. What writers like Prinz ought to be doing instead is revising their philosophical intuitions in light of developments in the empirical sciences.

My suggestion in this thesis is that an objective interpretation of recent neural evidence indicates that traditional beliefs about the nature of emotion are fundamentally mistaken. This evidence reveals that emotion is not reducible to either the perception of a somatic state or a cognitive appraisal of one's overall situation alone. Rather, it suggests that emotional experience is generated by a complex interplay of cognitive appraisals, sensory processes, and representations in our working memory. Reconstructing emotional consciousness in this manner, I have argued, could have a profound influence on philosophical and psychological debates about the fundamental nature of moral judgment. In fact, an integrative account of emotional consciousness could potentially bridge the gap between rationalist theories and sentimentalist theories of moral judgment.

My proposal, however, is not without its own difficulties. First, I would like to clarify that I do not think that neuroscience alone can be relied on to solve philosophical problems. Rather, it should be used as a resource for developing empirically-informed

philosophical theories. This applies not only to theories of moral judgment, but to other philosophical problems as well.

Moreover, while great progress has recently been made in the field of cognitive neuroscience, we must remember that many neuroscientific theories are themselves speculative. Given that neural researchers are continually improving the methods and the apparatuses they use to study the brain, it is certain that contemporary neural theories will likewise improve and evolve in the future. When such change does occur, we must also be prepared to rethink our philosophical ideas accordingly.

What my thesis has shown, most importantly, is how allowing alternative disciplines to influence long-standing philosophical debates can generate fresh ideas and encourage the development of new theories. This is not to say that our traditional philosophical views about issues like moral judgment are fundamentally mistaken. I simply wish to emphasize that it is often when we view philosophical problems from an alternative perspective that progress becomes possible.

References

- Adolphs, R. "Cognitive neuroscience of human social behaviour." *Nature Reviews: Neuroscience* 4 (2003): 165-178.
- Aubie, B. and P. Thagard. "Coherence in the brain: A neurocomputational model of parallel constraint satisfaction." Forthcoming-a.
- Aubie, B. and P. Thagard. "Emotional appraisal: A distributed neurocomputational model." Forthcoming-b.
- Ayer, A.J. *Language, Truth, and Logic*. New York: Dover, 1956.
- Ben Ze'ev, Aaron. *The Subtlety of Emotions*. Cambridge: MIT Press, 2000.
- Berthoz, S., et al. "An fMRI study of intentional and unintentional (embarrassing) violations of social norms." *Brain* 125 (2002): 1696-708.
- Blackburn, Simon. *Essays in Quasi-Realism*. New York: Oxford University Press, 1994.
- Blackburn, Simon. *Ruling Passions*. Oxford: Clarendon Press, 1998.
- Blair, R. "A cognitive developmental approach to morality: Investigating the psychopath." *Cognition* 57 (1995): 1-29.
- Blair, R., et al. "A selective impairment in the processing of sad and fearful expressions in children with psychopathic tendencies." *Journal of Abnormal Child Psychology* 29 (2001): 491-98.
- Blair, R., et al. "The psychopathic individual: A lack of responsiveness to distress cues?" *Psychophysiology* 34 (1997): 192-98.
- Blair, R., et al. "Turning a deaf ear to fear: Impaired recognition of vocal affect in psychopathic individuals." *Journal of Abnormal Psychology* 111 (2002): 682-86.
- Bowden, Peta. *Caring: Gender Sensitive Ethics*. London: Routledge, 1997.
- Casebeer, W. and P.S. Churchland. "The Neural mechanisms of Moral Cognition: A Multiple-Aspect Approach to Moral Judgment and Decision-Making." *Biology and Philosophy* 18 (2003): 169-94.
- Damasio, Antonio. *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Penguin Books, 1994.
- D'arms, J. and D. Jacobson. "Sentiment and Value." *Ethics* 110 (2000): 722-48.

- Darwall, S., A. Gibbard, and P. Railton. *Moral Discourse and Practice*. New York: Oxford University Press, 1997.
- Diogenes Laertius. *Lives of Eminent Philosophers*. Vol. 2. Trans. R.D. Hicks. Cambridge: Harvard University Press, 1925.
- Eliasmith, C. and C.H. Anderson. *Neural Engineering: Computation, representation and dynamics in neurobiological systems*. Cambridge: MIT Press, 2003.
- Gibbard, Allan. *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Cambridge: Harvard University Press, 1990.
- Gilligan, Carol. *In a Different Voice*. Cambridge: Harvard University Press, 1982.
- Greene, Joshua and Jonathan Haidt. "How (and where) does moral judgment work?" *Trends in Cognitive Neuroscience* 6.12 (2002): 517-523.
- Greene, Joshua, et al. "An fMRI investigation of emotional engagement in moral judgment." *Science* 293 (2001): 2105-08.
- Griffiths, Paul. *What Emotions Really Are: The Problem of Psychological Categories*. Chicago: University of Chicago Press, 1997.
- Hume, David. *Treatise of Human Nature*. Ed. L.A. Selby-Bigge. Oxford: Clarendon Press, 1967.
- James, W. and C. Lange. *Emotions*. New York: Hafner Publishing Company, 1922.
- Kant, Immanuel. *Grounding for the Metaphysics of Morals*. 3rd ed. Trans. James Ellington. Indianapolis: Hackett, 1993.
- Kaplan, J., J. Freedman, and M. Iacoboni. "Us vs. them: Political attitudes and party affiliation influence neural response to faces of presidential candidates." Forthcoming.
- Kohlberg, Lawrence. "From Is to Ought: How to commit the naturalistic fallacy and get away with it in the study of moral development." *Cognitive Development and Epistemology*. Ed. T. Mischel. New York: Academic Press, 1971.
- Lazarus, R. and B. Lazarus. *Passion and Reason: Making Sense of Our Emotions*. New York: Oxford University Press, 1994.
- Litt, A., C. Eliasmith, and P. Thagard. "Why losses loom larger than gains: Modeling neural mechanisms of cognitive-affective interaction." *Proceedings of the twenty-eighth annual meeting of the Cognitive Science Society*. Ed. R. Sun and N. Miyake. Mahwah: Erlbaum, 2006.

- Litt, A., C. Eliasmith, and P. Thagard. "Neural affective decision theory: Choices, brains, and emotions." Forthcoming.
- McDowell, J. *Mind and World*. Cambridge: Harvard University Press, 1994
- McDowell, J. *Mind, Value, and Reality*. Cambridge: Harvard University Press, 1998.
- Mill, J.S. *Utilitarianism*. 2nd ed. Ed. George Sher. Indianapolis: Hackett, 2001.
- Miller, Alexander. *An Introduction to Contemporary Metaethics*. Cambridge: Polity, 2003.
- Moll, J., R. De Oliveira-Sousa, and P. Eslinger. "Morals and the human brain: A working model." *Neuroreport* 14 (2003): 299-305.
- Morris, J.S. "How do you feel?" *Trends in Cognitive Sciences* 6 (2002): 317-19.
- Murphy, S., J. Haidt, and F. Björklund. "Moral dumbfounding: When intuition finds no reason." Forthcoming.
- Nichols, Shaun. *Sentimental Rules: on the natural foundations of moral judgment*. Oxford: Oxford University Press, 2004.
- Nichols, Shaun. "Sentimentalism naturalised." *The psychology and biology of morality*. Ed. W. Sinnott-Armstrong. Forthcoming.
- Nerb, J. "The dynamics of the appraisal-emotion relationship." *Cognition and Emotion*. Forthcoming.
- Noddings, Nel. *Caring: A Feminine Approach to Ethics and Moral Education*. Berkeley: University of California Press, 1984.
- Nussbaum, Martha. *Upheavals of Thought: The Intelligence of Emotions*. Cambridge: Cambridge University Press, 2001.
- Oatley, K. *Best Laid Schemes: The Psychology of Emotions*. Cambridge: Cambridge University Press, 1992.
- Oakley, K. *Emotions: A Brief History*. Malden: Blackwell, 2004.
- Ortony, A., G. Clore, and A. Collins. *The Cognitive Structure of Emotions*. Cambridge: Cambridge University Press, 1988.
- Piaget, Jean. *The Moral Judgment of the Child*. Trans. Marjorie Gabain. Glencoe: Free Press, 1948.

- Plato. *The Republic*. Trans. Allan Bloom. New York: Basic Books, 1991.
- Plato. *Phaedrus*. Trans. Christopher Rowe. London: Penguin Books, 2005.
- Prinz, Jesse. *Gut Reactions: A Perceptual Theory of Emotion*. New York: Oxford University Press, 2004.
- Prinz, Jesse. "The Emotional Basis of Moral Judgments." *Philosophical Explorations* 9.1 (2006): 29-43.
- Railton, Peter. *Facts, Values, and Norms: Essays Toward a Morality of Consequence*. Cambridge: Cambridge University Press, 2003.
- Rawls, John. *A Theory of Justice*. Cambridge: Harvard University Press, 1972.
- Rolls, E.R. *Emotion Explained*. Oxford: Oxford University Press, 2005.
- Sanfey, et al. "The neural basis of economic decision-making in the ultimatum game." *Science* 300: 1755-58.
- Scherer, K. *Facets of Emotion*. Hillsdale: Lawrence Erlbaum Associates, 1988.
- Schnall, S., J. Haidt, and G. Clore. "Irrelevant disgust makes moral judgment more severe, for those who listen to their bodies." Forthcoming.
- Shweder, R. et al. "The 'big three' of morality (autonomy, community, and divinity), and the 'big three' explanations in suffering." *Morality and health*. Ed. A. Brandt and P. Rozin. New York: Routledge, 1997.
- Smith, Adam. *The Theory of Moral Sentiments*. New York: Garland, 1971.
- Smith, E. and J. Jonides. "Storage and executive processes in the frontal lobes." *Science* 283 (1999): 1657-61.
- Solomon, R. *In Defense of Sentimentality*. New York: Oxford University Press, 2004.
- Solomon, R. *Not Passion's Slave*. New York: Oxford University Press, 2003.
- Solomon, R. *The Passions*. New York: Anchor Press, 1976.
- Thagard, P. "Emotional Consciousness: A neurocomputational theory." Forthcoming.
- Wagar, B.M. and P. Thagard. "Spiking Phineas Gage: A neurocomputational theory of cognitive-affective integration in decision making." *Psychological Review* 111 (2004): 67-79.

Wheatley, T. and J. Haidt. "Hypnotically induced disgust makes moral judgments more severe." *Psychological Science* 16 (2005): 780-84.

Wicker, et al. "Both of Us Disgusted in My Insula: The Common Neural Basis of Seeing and Feeling Disgust." *Neuron* 40 (2003): 655-64.

Wiggins, David. "A Sensible Subjectivism." *Needs, values, truth: Essays in the philosophy of value*. Oxford: Blackwell, 1987.