10-18-2016

Neo-sentimentalism and the bodily attitudinal theory of emotions

Chun Nam, Emile CHAN

Terms of Use

NEO-SENTIMENTALISM
AND
THE BODILY ATTITUDINAL THEORY OF EMOTIONS

CHAN CHUN NAM EMILE

MPHIL

LINGNAN UNIVERSITY

2016

NEO-SENTIMENTALISM
AND
THE BODILY ATTITUDINAL THEORY OF EMOTIONS

by

Chan Chun Nam Emile

陳鎮南

A thesis
submitted in partial fulfillment
of the requirements for the Degree of
Master of Philosophy in Philosophy

Lingnan University

2016

ABSTRACT

Neo-sentimentalism
and
The Bodily Attitudinal Theory of Emotions

by

Chan Chun Nam Emile

Master of Philosophy

Section 1 of this thesis investigates one issue in meta-ethics, namely, the nature of moral judgments. What are moral judgments? What does it mean by "wrong" when we assert "Killing is wrong?" Neo-sentimentalism is a meta-ethical theory which holds that the judgment that killing wrong is the judgment that it is appropriate to have a particular negative emotion towards the action. In other words, to judge that murder is wrong is to judge that we have a right reason for having a negative emotion towards the behavior. In the framework of neo-sentimentalism, the concepts of wrongness consist of negative emotions.

If the moral judgment is the judgment that it is appropriate to have a negative emotion towards the action, and the concept of wrongness contains a negative emotion, then the following question is what emotions are. In section 2, I endorse the bodily attitudinal theory of emotions, a view which holds that a conscious physiological reaction which induces behavior disposition and the change of facial expression and internal organs is necessary for having an emotion. This section also articulates and replies to three major objections towards the bodily attitudinal theory of emotion.

DECLARATION

I declare that this is an original work based primarily on my own research, and I warrant that all citations of previous research, published or unpublished, have been duly acknowledged.

_____

(CHAN CHUN NAM Emile)

CERTIFICATE OF APPROVAL OF THESIS

NEO-SENTIMENTALISM
AND
THE BODILY ATTITUDINAL THOERY OF EMOTIONS
by
CHAN Chun Nam Emile

Master of Philosophy

Panel of Examiners :

(Prof. HO Hung Lam Elizabeth) (Chairman)

(Prof. WONG Kai Yee) (External Member)

(Prof. NADO Jennifer) (Internal Member)

(Prof. BAKER Derek) (Internal Member)

Chief Supervisor

Prof. BAKER Derek

Co-supervisor

Prof. LIVINGSTON Paisley

Approved for the Senate :

(Prof. MOK Ka Ho Joshua)
Chairman, Postgraduate Studies Committee

18 OCT 2016

Date

CONTENTS

Acknowledgements

Thanks to my supervisors for providing extensive linguistic and stylistic corrections to the earlier drafts of this thesis. The remaining errors are my own.

I sincerely thank Prof. Derek Clayton Baker and Prof. Paisley Nathan Livingston. What they have taught me are more than philosophy.

0 Overview of the thesis

This thesis investigates an issue in meta-ethics, namely, the nature of moral judgments. Before giving an answer to this issue, it would be better to explain what meta-ethics is. Fisher (2011, 11) mentions an analogy to distinguish between applied ethics, normative ethics, and meta-ethics. Imagine that studying ethics is just like playing football. In the game, there are three groups of people, including players, referees, and analysts. Applied ethics are like players, who reflect that whether their actions violate the given rules. They study in issues like if it is morally wrong to have an abortion and whether it is morally required to give money to charity according to the given moral norms. Normative ethicists are like referees, who interpret the rules that govern the game. They investigate on questions like whether consequences matter to decide the moral status of having an abortion or giving money to charity. Meta-ethics are like analysts, who attempt to understand how the football game goes on. They explore questions like what it means by "morally praiseworthy" and "morally blameworthy" and so forth, how to know what is right and what is wrong, and how to be motivated to act morally.

As I have briefly distinguished between applied ethics, normative ethics, and meta-ethics, it is time to ask what the nature of moral judgments is. In this thesis, I attempt to propose moral sentimentalism, briefly, a view which holds that emotions are necessary for moral judgments. The sentimentalist theory possesses some advantages. For example, it explains why moral judgments are motivating because emotions are motivating and indispensable to moral judgments. Also, our daily experience tells us that most of the time we will feel anger when we judge that one has acted morally wrong, or we will feel guilt when we judge that we have done something immoral.

Historically, David Hume was one of the pioneers who proposed moral sentimentalism, a view that moral judgments depend on the feeling of sentient beings. As in his work *An Enquiry concerning the Principles of Morals*, David Hume writes:

> The final sentence, it is probable, which pronounces characters and actions amiable or odious, praise-worthy or blameable; that which renders morality and active principle, and constitutes virtue our happiness, and vice our misery; It is probable, I say, that this final sentence depends on some internal sense of feeling, which nature has made universal in the whole species. (Beauchamp ed., 1998, 5)

Such feeling of sentient beings is what Hume called *sympathy* which we refer to as *empathy* nowadays (Slote, 2010, 5). Following Hume's sprit, Slote develops a particular version of moral sentimentalism. Slote (2010, 63; Watson, 2011 143) understands moral judgments as moral approvals (or disapprovals) that are analyzed in terms of the feelings of being warmed (or being chilled) by one's empathy towards others' pain or needs.

However, sometimes we still judge an action as morally wrong while being warmed by one's empathy towards another person (Watson, 2011, 143). It leads us to believe that it is needed to have a less specific version of moral sentimentalism in order to avoid the counter-example that Slote's sentimentalism suffers from. A simpler version of moral sentimentalism is proposed by Jesse Prinz. As he writes:

> The theory I have in mind is not new. It's a variant of an old theme, associated with the British moralists, especially Hume. Simply, the theory says: To believe that something is morally wrong (right) is to have a sentiment of disapprobation (appropriation) towards it. (Prinz, 2006, 33)

Prinz's sentimentalism is similar to D'Arms and Jacobson's (2000, 724) sentimentalist account which holds that "the judgment that an action X is wrong with disapproval of X (and would similarly identify rightness with approval)."

Prinz's theory takes negative sentiments in general, instead of empathy specifically, as what are necessary to moral judgments, and thus avoid the counter-examples to Slote's sentimentalism. Yet such simple form of moral sentimentalism faces a direct objection,

namely, the problem of the recalcitrance of emotions. Sometimes we judge that an action has a negative value without having a negative sentiment towards the action, or we have a negative sentiment towards the action without judging the action as having the negative value (D'arms and Jacobson, 2000, 725). It is needed to modify the simple form of moral sentimentalism to overcome this objection.

The next question is how to modify the simple sentimentalism. By looking back to what Hume writes, a glue can be found:

> But in order to pave the way for such a sentiment, and give a proper discernment of its object, it is often necessary, we find, that much reasoning should precede, that nice distinctions be made, just conclusions drawn, distant comparisons formed, complicated relations examined, and general facts fixed and ascertained. (Beauchamp ed., 1998, 5)

As D'Arms and Jacobson (2000, 723) also suggest, "hence, philosophers who take inspiration from Hume must allow reasoning, as well as feeling, to play a role in evaluative judgment. The central challenge for sentimentalism is to preserve the idea that values are somehow grounded in the sentiments, while at the same time making sense of the rational aspects of evaluation." On the basis of what Hume and D'Arms and Jacobson mention, it is natural and legitimate to add reason into the sentimentalist framework to avoid the problem that the simple sentimentalism suffers from.

Neo-sentimentalism is a theory that unifies emotions and reason, as Sauer writes:

> To judge an action A to have some evaluative property P (goodness, badness, rightness, wrongness…) is to judge it appropriate (warranted, justified…) to have an associated emotional response (guilt, resentment, contempt…) towards A. (Sauer, 2011, 111)

It is said that to judge that an action is morally wrong is to judge that it is appropriate to have a negative sentiment, such as anger, towards the action. But it does not mean that the judgment of an appropriate emotion towards the action is actually accompanied by a negative emotion. Otherwise, neo-sentimentalism cannot be immune to the problem of

recalcitrant of emotions. Because it is obvious that one can judge that it is appropriate to have a negative emotion without having the emotions actually.

Nevertheless, once we understand moral judgments as the judgments of appropriate emotions without actually having the emotions, what is the importance of emotions? It seems that emotions no longer a play crucial in the picture. To preserve the importance of emotions, I introduce the dual-process theory of mind. The theory holds that the mind is divided into two systems, namely, system 1 and system 2. System1 is responsible for automatic and instinctive processes that do not need to be consciously accessed and are cognitively economical, whereas system 2 is responsible for rational and deliberative processes that require conscious attentions and are cognitive-resources demanding. I shall argue that emotions belong to system 1 and thus one makes moral evaluations cognitively economical. In addition, emotions make one sensitive to the values of the objects in environments and motivate one to act according to the moral judgment. Without the help of emotions, one cannot be motivated to act morally.

Since emotions have such importance, I need to introduce a theory of emotions to illustrate their significance. We need an account of emotions that illuminates how emotions make one aware of the value of objects, able to make implicit moral evaluations, and motivated to act morally. James's theory of emotions was once influential. His theory holds that the changes of the internal bodily feelings are emotions themselves. As James writes:

> My thesis, on the contrary, is that the bodily changes follow directly the perception of the exciting fact and that our feeling of the same changes as they occur is the emotion. James (1884, 189)

Nonetheless, understand emotions as what James suggests leads us to a serious flaw. Our ordinary speaking of emotions demonstrates that emotions are intentional, which means that emotions are directed at objects in the world. For instance, "Sam is sad for his cat, and "Mary is angry at the man" and so forth. If emotions are merely the feelings of the internal

bodily changes, then James's theory cannot explain the intentionality of emotions. And as James's theory cannot capture the intentionality, his theory cannot demonstrate how emotions make one sensitive to the values of objects. The theory only makes one aware of the feelings of one's internal bodily changes.

The bodily attitudinal theory of emotions can be considered as a descendant of James's theory. Yet unlike its ascendant, the bodily attitudinal theory not only captures the phenomenology but also the intentionality of emotions, as the following is its central thesis:

> Each emotion consists in a specific felt bodily stance towards objects or situations, which is correct or incorrect as a function of whether or not these objects and situations exemplify the relevant evaluative property. (Deonna and Teroni, 2012, 89)

The bodily attitudinal theory of emotions holds that to have an emotion is to have a bodily attitude towards the object. The bodily attitude is a consciously felt physiological response to the object, which consists of "action readiness," the feelings of the internal bodily changes and other bodily phenomena. Action readiness is an idea proposed by the psychologist Nico Frijda. The idea is that to have action readiness is to have the behavioral tendency to perform a particular kind of behavior. To fear the dog is to have a bodily attitude towards the dog and to have the attitude is to feel your body as poised to escape from the dog, with the feelings of your internal bodily changes and other bodily phenomena.

If we understand emotions as a consciously felt physiological response towards objects, then we can illustrate the importance of emotions. Emotions make us sensitive to the values of objects because of the bodily attitudes that are consciously felt. Emotions motivate us to act morally because to have a bodily attitude is to feel our body as poised to act in a certain way. And such behavioral dispositions are implicit evaluations of objects.

Although the bodily attitudinal theory of emotions illustrates the importance of emotions, the theory faces at least one serious difficulty. It is said that some patients who suffer from severe spinal cord damage reported that they experience conscious emotions, (Smith, 2014). It is supposed that the patients are unable to feel the feelings of their internal bodily changes and perform the muscular movements since their brain is disconnected or poorly connected to their body. Thus, it seems that bodily attitudes are not necessary for emotions, which means that the bodily attitudinal theory encounters counter-examples.

To protect the bodily attitudinal theory, my contribution is to introduce the idea of "phantom emotions" to explain why those patients are still able to have consciously emotions. My idea of phantom emotions is inspired by the occurrence of phantom limbs. The majority of amputees reported that they felt their limbs even internal organs that had been removed. They could vividly feel the shape, posture, and movement of their moved limbs and organs. On the basis of the occurrence of phantom limbs, I attempt to use phantom emotions, as a speculation, to explain why those patients can still experience conscious emotions since bodily attitudes consist of action readiness and other bodily phenomena that could emerge as phantoms.

This thesis is divided into two sections. Section 1 investigates the nature of moral judgments. In section 1.1 I shall mention empirical evidence that supports the link between emotions and moral judgments. In section 1.2 I articulate different versions of moral sentimentalism and propose neo-sentimentalism dialectically, in order to provide a conceptual account that explains the nature of moral judgments that emotions have a role within. Since according to neo-sentimentalism, moral judgments are analyzed in terms of the judgments of an appropriate emotion, it is needed to clarify the notion of appropriateness. In section 1.3 I shall propose three

prominent theories of reasons of the right kind as possible alternatives to clarify the notion of being appropriate. The neo-sentimentalist understands moral judgments as the judgments of an appropriate emotion without actually having the emotions, which means that the significance of emotions seems largely weakened in the picture. To preserve the importance of emotions, in section 1.4 I shall introduce the dual-process theory to argue that emotions are a more cognitively economical way to make moral evaluations.

To demonstrate other importance of emotions, namely, how emotions make us sensitive to the value of objects and motivate us to act morally, in section 2 I shall introduce the bodily attitudinal theory of emotions. Section 2.1 clarify the central thesis of the bodily attitudinal theory of emotions and explain why we should be in favor the bodily attitudinal theory of emotions rather than its rival, the judgmental theory of emotions. Section 2.2 is a discussion about how to justify emotions and value judgments. Section 2.3 explain the intentionality of emotions in the framework of the bodily attitudinal theory by introducing the teleological theory of representation. Section 2.4 considers and replies to possible objections to the bodily attitudinal theory. And section 2.5 proposes a method to derived moral emotions from non-moral emotions.

1 Moral Sentimentalism

Section 1 investigates moral sentimentalism. Briefly, moral sentimentalism is a meta-ethics theory which holds that emotions are necessary to moral judgments. This section is divided into four sub-sections. Section 1.1 explores empirical evidence which supports the link between emotions and moral judgments. Section

1.2 articulate different versions of moral sentimentalism. Section 1.3 examines the relation between the notion of appropriateness and reasons of the right kind. Section 1.4 replies to objections to neo-sentimentalism once the idea of appropriateness has been introduced into the theory.

1.1 Empirical Evidence for the Link Between Emotions and Moral Judgments

Before endorsing his version of moral sentimentalism, Prinz (2006, 30-33; 2007; 21-47) gather a fruitful amount of empirical evidence which supports the thesis that emotions are related to moral judgments. Prinz understands the relation in terms of the conjunction of three theses. The first thesis is that emotions co-occur with moral judgments. The second thesis is that emotions affect moral judgment. And the third thesis is that emotions are necessary for moral judgments.

1.1.1 Emotions Co-occur with Moral Judgments

Oliveira-Souza and Eslinger (2003, 301) provides the first piece of evidence to support this thesis. While measuring the subjects' brain activities, they invited the subjects to evaluate moral sentence such as "you should break the law when necessary" and non-moral sentence such as "stones are made of water." To answer both kinds of sentences, the subjects only answered "right" or "wrong." The emotional areas of the subject's brain were active while evaluating moral sentences whereas the areas were not active while evaluating non-moral judgments.

Sanfey et al. (2003, 1757) provide another piece of evidence to support this thesis. Two subjects were asked to play the ultimate game, a game which involves distributing monetary resources, Sanfey and his colleagues measured the subject's

brain activity during the play. While playing the game, one subject divided the monetary sums with another player. If another player found that the distribution of resources was unfair, the emotional areas of the player's brain would be activated.

Berthoz et al. (2002, 1698) collected the third piece of evidence to support the thesis. Berthoz and the others mentioned a story to the subjects. The story is that one dinner guest impolitely spat food back onto the plate after the first bite and judge the food revolting. The emotions areas of the subject's brain were activated, and the subject considered the behavior to be an inappropriate violation of a norm.

1.1.2 Emotion Affect Moral Judgments

Prinz (2006, 31) mentions how other psychologists support the thesis that emotions affect moral judgments. Schnall and the others found that if the subjects had a more negative emotion, a more negative moral judgment would be made. They mentioned the following scenario to the subjects:

> Frank's dog was killed by a car in front of his house. Frank had heard that in China people occasionally eat dog meat, and he was curious what it tasted like. So he cut up the body and cooked it and ate it for dinner. How wrong is it for Frank to eat his dead dog for dinner? (Schnall et al., 2008, 16)

The subjects were asked to answer a questionnaire about the scenario and were divided into two groups. The subjects in the first group gave their answers on a clean and neat table which unlikely caused disgust, whereas the subjects in the second group gave their answers on a nasty table with a dirty cup, a broken pencil, used tissues, and a greasy pizza box which likely caused disgust. The subjects in the second group judged the action as morally worse whereas the subjects in the first

group judged the action less morally wrong. Schnall and the others interpreted this piece of evidence as suggesting that emotions affect moral judgments.

However, this interpretation of proof can be challenged. It might be the case that the subjects' attentions were drawn to situations which possess relevant moral properties, which means that the subject associated the scenario with other actions or conditions which we deem morally questionable. It needs to establish that negative emotions can solely make us have “negative moral judgments even when we have no reason” for believing that the scenario does not possess moral wrongness (Prinz, 2006, 31).

Wheatley and Haidt (2005, 780) found a piece of evidence to support the needed proof. Wheatley, Haidt, and the others hypnotized subjects to make them experience disgust suddenly when listening to the word "often." the subjects were told stories which contain the word "often" and other synonymous. After being hypnotized, the subjects felt disgusted when hearing the word "often" in the stories. Interestingly, although some of the stories containing many tokens of the word "often" described morally admirable characteristics, the subjects judged those traits as morally condemnable. This piece of evidence leads to the conclusion that negative emotions can solely cause us to have an adverse moral judgment even when we would otherwise find nothing morally objectionable in the scenarios.

Besides, Murphy et.al. (2000, 18) invited subjects to evaluate the stories and questions given in their experiments. One of the stories describes two siblings having sexual intercourse. Björklund and the other asked the subjects to justify why the sibling incest is morally unacceptable and denied the subject's justifications. For

example, the subjects claimed the behavior is morally wrong because they will give birth to children with congenital defects, but Björklund and the others told the subjects that birth control is applied. After the subjects' justification had been challenged in this manner, the majority of the subjects still thought the action immoral because it is disgusting, whereas only a handful of subjects changed their minds about the moral status of incest (Prinz, 2006, 31). Assuming there are no other possibilities, either reason or emotions affect moral judgments. If the subjects still believed that sibling incest is morally problematic after their justifications had been refuted, then it seems that emotions affect moral judgments.

1.1.3 Emotions Are Necessary for Moral Judgments

The claim that emotions are necessary for moral judgments is stronger than the claim that emotions co-occur with moral judgments. The former suggests that emotions are necessary to moral judgments, which means that emotions are indispensable to moral judgments. Instead, the latter merely assets that emotions and moral judgments emerge in the same situation at the same time. In this sense, emotions may not be part of moral judgments.

The existence of psychopathy provides strong support to the thesis that emotions are necessary for moral judgment (Prinz, 2006, 32). Psychopaths are those who have significant difficulties with experiencing pain and negative emotions like fear and sadness. It is also hard for psychopaths to identify facial expressions and tones which are related to negative emotions (Blair et al., 2001, 491-492). In other words, psychopaths are significantly incapable of feeling negative emotions in themselves and detecting negative emotions in others. Because of this incapacity,

psychopaths keep committing morally unacceptable actions. Prinz (2006, 32) believes that psychopaths act immorally because they are unable to have genuine moral judgments, which means that psychopaths pretend that they understand what "morally right" and "morally wrong" mean, but in fact, psychopaths cannot comprehend moral concepts in the way emotionally healthy people do.

Blair (1995, 25) proposed that psychopaths are unable to differentiate between ethical and conventional rules. If we give psychopaths a set of rules and ask them which ones are moral and which ones are conventional, psychopaths likely answer the questions incorrectly. Blair (1995, 11) provides a conjecture to explain why psychopaths are unable to distinguish between moral and conventional rules. The speculation is that psychopaths lack a violence-inhibition mechanism (VIM). In other words, psychopaths have a psychological problem which makes them unable to stop themselves from committing violence. The lack of VIM is explained by the lack of moral emotions, and the inability to sense others' negative emotions like fear and sadness.

Nonetheless, Prinz (2007, 45) disagrees with that the VIM model can satisfactorily explain psychopathy in the three following way. Firstly, psychopaths also commit many morally condemnable actions such as lying and manipulating which do not involve violence. Secondly, the VIM model cannot explain psychopathy's cognitive deficiency. For instance, psychopaths make numerous mistakes while finishing mazes with increasing difficulties. Thirdly, the VIM model does not explain why psychopaths lack non-moral negative emotions like fear and

sadness besides the lack of moral emotions such as empathy. Another model is needed to explain psychopathic behavior.

Prinz (2007, 46) believes that Gray's Behavior Inhibition System (BIS) is more capable of explaining psychopathic behavior since the core of the model is made of non-moral negative emotions such as anxiety and grief and negative moral emotions like empathic distress. The lack of the mechanism also brings psychopaths the deficit with changing or ceasing their behavior.

1.1.4 Emotions, Moral judgments, and Motivational Internalism

If emotions are necessary for moral judgments, then motivational internalism is sustained. The strong version of motivational internalism holds that one's genuine moral judgment brings one a dominant motivation to act accordingly. To illustrate, if one judges that smoking is morally impermissible, then one is overridingly motivated not to smoke even if the temptation to smoke is present. Still, this version of motivational internalism is unreasonably strong because the theory leaves no room for the weakness of will and inclinations against the moral behavior. The weak version of motivational internalism holds that one's sincere ethical judgment brings one a defensible motivation to act accordingly, which means the moral motivation can be overridden by conflicting desires or mental illness (Rosati, 2014, section 3.2).

Prinz (2015, 70) mentions an argument for motivational internalism. Although Prinz does not say which version of motivational internalism is supported by his argument; I assume that Prinz endorses the weak version of motivational internalism for the sake of being conservative. Prinz's argument is that moral

judgments consist of emotional attitudes, and emotional attitudes are motivating. Therefore, moral judgments are motivating. Prinz explicitly uses the phrase "emotional attitudes" rather than "emotions" because emotions are directed at objects and they are evaluations on objects. I shall clarify what attitudes mean in section 2.

Since the premise that moral judgments contain emotions is sustained by the empirical evidence mentioned in the previous sub-sections, I focus on justifying the claim that emotions are motivating. Although our daily experience tells us that emotions are motivating, the claim will be more fortified if empirical evidence is brought to support it. McMillen and Austin (1971, 60) suggest that moral emotions play a major role in promoting pro-social behavior. Evidence indicates that one who feels guilty is more willing to perform compensational actions. For instance, subjects who had cheated in an examination helped people for 62 minutes averagely, while those who had not cheated in the test helped people merely for 2 minutes. Feeling guilty makes one bring compensation, and anticipating the feeling makes one abler to resist future temptations. Furthermore, Prinz and Nichols (2010, 130) studied the psychological profile of anger. If a person has done a wrongdoing to one, the one will feel angry at the person and are motivated to retaliate. Intense anger would even motivate one to seek vengeance with a risk of paying a heavy price. These studies suggest that emotions are motivating.

1.2 Different Versions of Moral Sentimentalism

The previous sub-section mentions how emotions are related to moral judgments, and we have reasons for believing that moral judgment consists of

emotions. Then the following question is: Emotions are necessary to moral judgments in what way? I shall discuss Slote's and Prinz's moral sentimentalism and neo-sentimentalism below.

1.2.1 Slote's Sentimentalism

To introduce and articulate the version of moral sentimentalism which I endorse, I shall first describe Michael Slote's moral sentimentalism because his version is specific and one of the strongest accounts of moral sentimentalism. The central thesis of Slote's account is as follows:

> (SS) An action is judged morally admirable (or morally condemnable) if and only if the action manifests one's agential empathy (or the lack of agential empathy), and we feel warmed (or chilled) by one's agential empathy (or the lack of agential empathy). (Slote, 2010, 63; Watson, 2011, 143)

Moral judgments in (SS) consist of two layers. The first layer involves one's agential empathy towards another person. When one shows one's agential empathy towards another person, one shows an emotional response to the person's pain or needs. Clearly speaking, as Slote and Watson understand it, one' agential empathy is that one feels what another person feels:

> I pointed out earlier: namely, that sort of "sympathy" Hume is talking about is or involves what we nowadays usually refer to as empathy (rather than as sympathy) the kind of phenomenon Bill Clinton was invoking or referring to when he said, "I feel your pain." (Slote, 2010, 30)
>
> When I say I empathize, I do mean "empathize" and not "sympathize," I feel for her pain. I can vividly imagine what it would be like to be her… (Watson, 2011, 146)

The second layer is where moral approval (or disapproval) comes out. This moral approval (or disapproval) is our empathy towards one's agential empathy. We are warmed if one demonstrates one's agential empathy towards another person, whereas we are chilled if one shows the lack of one's agential empathy towards another person. This empathetic warmth (or chill) is moral approval (or disapproval).

Watson (2011, 143) mentions a counter-example to (SS). Sometimes one demonstrates one's agential empathy towards another person, and we feel warmed by one's agential empathy, yet one's judged as morally wrong. We have a reason for rejecting (SS) if such counter-example exists:

> The example I wish to develop, as a test case, is of a battered woman who will not leave her husband, as is not uncommon. Imagine further that his refusal to leave is not purely practical or strategic but is grounded in the belief that he will change, that he does not really mean to harm her, that he really loves her, and so forth. To add details to the case, suppose her husband is a returning veteran, experiencing PTSD, and was a loving husband prior to combat. Yet he cannot get his violent outbursts under control. She feels conflicted and hopes that he will recover, but in the meantime, she suffers abuse. We think she is in a horrible circumstance, and it pains us to think of her plight. (Watson, 2011, 143)

We need to be aware of that in the above scenario there are two different tokens of agential empathy. The first token is the wife's empathy towards her husband which causes her choose to stay with her husband. The second token is our agential

empathy towards the wife as we think that she is a victim of such circumstance. The first token is what we should concern here.

It seems that we that although the wife's agential empathy towards her husband make us feel warmed, we tend to think that the wife is taking a morally wrong action because she chooses to stay with her husband. The right action would be to leave her husband to guarantee her safety and help her husband while keeping her distance.

To make sure that she treats (SS) fairly, Watson (2011, 144) think that we need to think whether the moral judgment that it is morally wrong for the wife to stay with her husband can be captured by (SS). Indeed, we are warmed by her agential empathy towards his husband, but we judge that her choice is morally wrong because she lacks agential empathy towards herself, and this lack chills us. If self-directed agential empathy is possible, then (SS) can avoid this counter-example. However, Slote explicitly states that self-care or self-directed action is not of our moral concern, and Slote only mentions that agential empathy is other-directed:

> Thus, empathy involves having the feelings of another (involuntarily) aroused in ourselves, as when we see another person in pain. (Slote, 2010, 15)

If Slote does not mention that empathy is not self-directed, then the above counter-example cannot be explained away by (SS).

1.2.2 Prinz's Sentimentalism

Prinz offers a less specific and strong account of moral sentimentalism. The following is the central thesis of Prinz's account:

> (PS) An action is judged morally wrong if and only if we have a negative sentiment towards the action. (Prinz, 2006, 33)

Prinz has his use of the word "sentiment." By "sentiment" Prinz means a disposition to have certain kinds of emotions:

> Sentiment of Approbation and disapprobation are, likewise, constituted by different emotions on different occasions." (Prinz, 2006, 34)

A disapprobation, in Prinz's understanding, is a category which involves negative emotions such as shame, guilt, anger, contempt, and disgust. If one deems what another person has done immoral, then one has the disposition to have a negative emotion, such as anger and disgust, towards another person's action. If one deems what one has done is morally condemnable, which means that one has the disposition to feel guilty or ashamed. Nonetheless, for the sake of simplicity, I put side the category issue and the words "sentiments" and "emotions" interchangeably here.

Although (PS) can avoid the counter-example to (PS) because (PS) does not take empathy as moral emotions in its framework, (PS) is still problematic since (PS) faces its counter-examples as well. There are two kinds of counter-examples which suggest that emotions are neither sufficient nor necessary to moral judgments. The first kind is the recalcitrance of emotions and the second kind is the recalcitrance of judgments.

The problem of the recalcitrance of emotions is used to attack the judgmental theory of emotions, a view which holds that value judgments are necessary to emotions. The problem is the judgmental theory preserves no room to

permit emotions to conflict with the corresponding value judgments (D'Arms and Jacabson, 2003, 129). This problem also bothers (PS). To have a recalcitrant emotion is to have the emotion even one does not believe the corresponding value judgment. The problem of the recalcitrance of judgments is that sometimes we have a value judgment even if the relevant emotion does not exist.

Interestingly, Prinz himself mentions the instantiations of the recalcitrance of emotions and judgments:

> Imagine a mob hit man who is going through a process of reform. He may not experience any sentiment of disapprobation towards killing, but he nevertheless judges that killing is wrong. That suggests that sentiments are not necessary. Or consider a person who was raised to think homosexuality is wrong, but now rejects that view. She may still harbor irrepressible negative sentiments when she thinks about homosexuality even though she insists that homosexuality is morally acceptable. That suggests that sentiments are not sufficient for moral judgments. (Prinz, 2007, 112)

Note that Prinz does not stand for a contradictory position because Prinz has held a consistent stance through years according to his works (Prinz, 2006, 33; 2007, 94, 2015, 70). Prinz mentions the above counter-examples for demonstrating how he could explain away the counter-examples by what he calls "second-order emotions."

1.2.3 Neo-Sentimentalism

The counter-example to (PS) leads us to believe that we need to adopt a more sophisticated version of moral sentimentalism, and neo-sentimentalism fits our purposes. The following is the central thesis of the theory:

> (NS) To judge an action A to have some evaluative property A (goodness, badness, rightness, wrongness…) is to judge it appropriate (warranted,

> justified…) to have an associated emotional response (guilt, resentment, contempt…) toward A. (Sauer, 2011, 111)

With the notion of appropriateness, (NS) can avoid the counter-examples to (PS). Prinz (2007, 114) illustrates how the idea of appropriateness can explain away the above counter-examples. Although the mob hit man experiences no negative emotion towards killing, he judges that killing is wrong because he deems feeling guilty for killing appropriate. Although the person experiences an irrepressible negative emotion towards homosexuality, she maintains that the judgment that being homosexual is morally permissible because she deems having a negative sentiment towards this sexual orientation inappropriate.

(NS) is immune to the counter-examples to (PS) because if it is appropriate to have a negative emotion towards killing, it means that we have a right reason for having the negative emotion, and the same reason is also for having the corresponding moral judgment. In other words, the concept of appropriateness prevents the recalcitrance of emotions and judgments by unifying emotions and moral judgment.

Prinz (2007, 113) attempts to analyze the notion of appropriateness as second-order emotions. A second order emotion is an emotion which is directed at a first-order emotion, an emotion directed at an object. The mob hit man deems having a negative emotion towards killing appropriate because he has a second-order negative emotion towards not having a negative emotion towards killing. The person deems having a negative emotion towards homosexuality inappropriate

because she has a second-order negative emotion towards having a negative emotion towards being homosexual.

Nonetheless, I disagree with that the notion of appropriateness can be analyzed in terms of second-order emotions. Firstly, it is unclear that all first-order emotions are accompanied by second-order emotions. Is my admiration towards the Mother Teresa's sacrifice accompanied by a second-order emotion? Secondly, assuming that second-order emotions accompany all first-order emotions, but what make second-order emotions appropriate? Third-order emotions? It seems that Prinz's analysis leads to an infinite regress. Perhaps some might say that it is not required to deem second-order emotions appropriate or inappropriate. Buy why not? Is it because second-order emotions are different from first-order emotions? But what is the difference? If first-order emotions need to be deemed appropriate or inappropriate, and there is no difference between first-order and second-order emotions, then it is unclear why it is not mandatory to deem second-order emotions appropriate or inappropriate as well. There are three other prominent ways of explaining the notion of appropriateness.

Furthermore, to judge that killing is morally wrong is to judge that it is appropriate to have a negative emotion towards killing, but we should not understand that the judgment of the appropriate emotion is always accompanied by an actual emotion. It is simply the judgment that it is appropriate to have the negative emotion. If the judgment of the appropriate emotion required an actual emotion, then neo-sentimentalism would suffer from the problem of the

recalcitrance of emotions as well, since it is possible that one judge that it is appropriate to resent killing without actually experiencing the anger towards killing.

1.3 Appropriateness and Reasons of the Right Kind

To be appropriate to have a negative emotion is to have a right reason for having the emotion. Based on the textual evidence, some scholars such Schroeder and Sauer, identify the notion of appropriateness with reasons of the right kind. "This is because the point of the distinction between the "right" and "wrong" kinds of reason, is that only the "right" kind contributes to standards of correctness…," Schroeder (2008, 13) writes. Sauer (2011, 115) repeatedly treat the notion of appropriateness the reasons of the right kind interchangeably. These pieces of textual evidence support that the concepts of appropriateness refer to reasons of the right kind.

1.3.1 Scanlon's Buck-Passing Account of Value

According to (NS), to judge that an action is wrong is to judge that it is appropriate to have a negative emotion towards the action, and this thesis is parallel to Scanlon's account:

> (BPV) The alternative, which I believe to be correct, is to hold that being good, or valuable, is not a property that itself provides a reason to respond to a thing in certain ways. Rather, to be good or valuable is to have other properties that constitute such reasons. (Scanlon, 1998, 97)

(BPV) holds that for an object to be valuable is not that the feature of being valuable "gives us a reason to respond to the object in a certain way" (Scanlon, 1998, 97), but is that the properties other than the property of being valuable to

give us a reason to show a particular response towards to the object. (Samuelsson, 2013, 1).

Since the role of giving us a reason for responding to the object in a certain way is passed to the properties other than the property of being valuable, the account is called a "buck-passing account" (Scanlon, 1998, 97). Applying (BPV) to (NS) amounts to saying that for an action to be morally wrong, is not a matter of the property of being morally wrong giving us a reason for responding to the behavior in a certain way. Instead, it is to say that the properties other the property of being morally wrong to give us a reason for having a particular negative emotional response towards the action.

1.3.2 Wrong Kinds of Reason Problem

Scanlon's account leads to wrong kinds of reason problem (WKRP). This issue holds that sometimes we have a reason for responding to an object in a certain way, yet the object itself lacks the corresponding value (or properties that constitute the value) (Rabinowicz and Rønnow-Rasmussen, 2004, 393; Samuelsson, 2013, 1). Rabinowicz and Rønnow (2004, 402) vividly illustrate this problem by inviting us to imagine that "a powerful demon threatens to impose severe pain on us uncles we admire the devil." This threat seems to give us a strong reason for praising the mighty fiend, yet we tend not to think that the demon is admirable (or has properties that constitute the properties of being admirable). However, according to (BPV), the demon's intention to impose severe pain on us is a property other than the property of being admirable gives us a reason for admiring the

demon. It leads us to believe that this reason for admiring the demon is a wrong reason, and thus, we need to exclude wrong reasons from right reasons (Samuelsson, 2013, 2).

Since the structure of (NS) is parallel to the framework (BPV), (NS) would suffer from the same problem if (WKRP) were not solved. I shall mention three prominent theories of reasons of the right kind which are alternatives to deal with (WKRP).

1.3.3 Lang and Samuelsson's Account of Reasons of the Right Kind

Lang proposes a theory of reasons of the right kind, and Lang's account is challenged by Olson's objection, then Samuelsson replies to Olson's criticism, which means that the account is co-developed by Lang and Samuelsson. Lang has modified (BPV) six times and the following is the central of the Lang's theory:

> (BPV6) An object is valuable if and only if the object has properties other than the property of being valuable that give us a reason to show a certain attitude towards the object, as long as those properties would still be reason-giving in the absence of *the benefits to us* of showing the certain attitude towards the object. (Lang, 2008, 484)

(BPV6) seems able to exclude the demon's threat from reasons of the right kind, because the threat as a property other than the property of being admirable, would not longer be reason-giving to us if the demon stopped threatening to punish us. In other words, if the demon does not force us to admire him, then we no longer have a reason for admiring the demon. A reason of the right kind must not be such a

prudential reason. Nonetheless, (BPV6) is further attacked by Olson's counter-example:

> Suppose for instance that an evil demon threatens to inflict severe pain on people on the other side of Earth, who are all strangers to us, unless we favour him. That seems to provide reason for us to favour the demon even though we would not benefit from favouring him. (Olson, 2009, 226)

Another wrong reason for admiring the devil comes into the picture. This counter-example is about the advantages of ours but of others. If the demon threatens to impose severe pain on people who we have never met, we seem to have a reason for admiring the devil, yet the demon still does not deserve being admired. To avoid the further counter-example and the other possible variations of the example, Samuelsson modifies (BPV6) in the following way:

> (BPV11) An object is valuable if and only if the object has properties other than the property of being valuable that give us a reason to show a certain attitude towards the object, as long as those properties would still be reason-giving in the absence of *any consequences* of showing the certain attitude towards the object. (Samuelsson, 2013, 7)

According to (BPV11), no matter what the demon would do to anyone, we would not have a right reason for admiring the devil, because any consequence that the devil would bring are not of our concern. (BPV11) is further objected because of its incompatibility with consequentialism, but I shall not discuss this criticism here.

1.3.4 Schroeder's Account of Reasons of the Right Kind

Before introducing his account of reasons of the right kind in his article, Schroeder distinguishes two kinds of wrong reasons. The first kind of wrong reasons

is that one engages in an activity for extra benefits. For instance, John admires the mighty devil for avoiding his punishment, or Jack believes the judgment that Mars is green for a huge amount of incentives. John and Jack's reasons are reasons for *some* people to engage in activities, but those reasons are not reasons for *anyone* to do something (Schroeder, 2008, 11).

Schroeder (2008, 12) illustrate the first kind of wrong reasons with the following example: if *your* son is being brought to court for murder, and *you* would experience a great torment if *you* believe that *your* son was guilty. Then *you* have a reason for refusing to accept the belief that *your* son is guilty, yet your reason is not anyone's reason for believing that your son is not guilty.

Schroeder (2008, 12) also mentions another example from Gregory Kavka. If a billionaire promises that he will give you a huge amount of incentives in exchange fro your intention to drink the poison *before tomorrow*, then you have a reason for intending to drink the position *now*, but this reason is only available at a *particular* time but *not* any time. The similarity shared by all of this two kinds of wrong reasons is that being *idiosyncratic*. By idiosyncratic, it means that these reasons are only valid for particular people who engage in certain activities. And the meaning of activities refers to any mental and physical actions in a broad sense.

The second kind of wrong reasons is that we sometimes conflate a kind of reasons with another kind. Schroeder believes that good instances of the second kind of wrong reasons are moral reasons. Imagine that someone makes a cruel but amusing joke at a gathering. We have a reason for not laughing at the joke because it is immoral to do so, but this moral reason does not make the joke less assuming

(D'Arm and Jacobson, 2000, 731). Moral reasons are universal in the way that the reasons are valid to anyone, whereas the idiosyncratic reasons are only valid to a particular kind of people who engage in a certain kind of activities.

After illustrating the characteristics of wrong reasons, Schroeder introduces an account to exclude such reasons. The following is the central thesis of Schroeder's theory:

> (SRKR) If a reason belongs to the right kind of reasons for engaging in an activity A if and only if the reason is shared necessarily by anyone who engages in A, and just because engaging in A (Schroeder, 2008, 13).

The "shared necessarily by anyone engages in A" part is important to differentiate the first kind of wrong reasons from right reasons, because the part is meant to exclude idiosyncratic reasons. If a reason is a right reason for engaging in an activity, then the reason is shared necessarily by anyone who engages in the activity and vice versa. The reason that the dog has properties that constitute fearsomeness is a right reason for fearing the dog because the reason is shared by necessarily who fears the dog. To illustrate, the dog performs an aggressive and impulsive behavior and has big claws and sharp teeth that are extremely likely to bring colossal even lethal damage to one's physical integrity. These features make any who has interactions with the dog believe that their physical integrity would be heavily damaged, or even their lives would be terminated if attacked by the dog. Thus, the features would constitute fearsomeness and are the reason for fearing dog that is shared by any who has interactions with the dog (if they are made of flesh and blood). Instead, the reason that a billionaire promises that he will give you a million

dollars if you fear the dog is not a right reason since the reason is not shared by necessarily who fears the dog. Some people simply are not interested in money or any benefits.

The "just because engaging in A" part is also crucial to differentiate the second kind of wrong reasons from right reasons because the part is meant to exclude universal reasons that can be applied to other activities. If a reason is a right reason for engaging in an activity, then the reason can only be applied the activity, and vice versa. The reason that the dog has properties that constitute fearsomeness is a reason for fearing the dog because this reason cannot be applied to other activities. On the contrary, a moral reason, say, God's command is not a right reason for fearing the dog because God's command can be applied to other activities.

1.3.5 Hieronymi's Account of Reasons of the Right Kind

According to Hieronymi (2005, 448; Birondo, 2014, 128), (WKRP) arises from our traditional understanding of reasons. Traditionally, we think that to have a reason is to have a consideration for an activity. The understanding conflates wrong reasons with right reason. If the demon threatens to impose severe pain on us unless we admire the demon, then we have a reason for admiring the devil, which means that we have a consideration for admiring the demon. Also, assume that the demon is in fact admirable (or has properties that constitute the property of being admirable), then we have another reason for admiring the demon, which means that we have another consideration for admiring the demon. Based on the traditional understanding of reasons, both reasons are also considerations for

admiring the demon, yet we have no resource for differentiating between these reasons which belong to different categories.

Therefore, Hieronymi suggests that we should abandon the traditional understanding and adopt her alternative:

> (HRKR) to have a right reason for engaging in an activity is to have a consideration which bears on the relevant question.

If the demon threatens to impose severe pain on us unless we admire the demon, then we have a reason for admiring the demon, which means that we have a consideration which bears on the prudential question whether it is beneficial to admire the demon. Assume that the demon is admirable (or has properties that constitute the property of being admirable), then we have a consideration for admiring the demon which bears on the emotional question whether the demon is admirable (or has properties that constitute the property of being admirable). For example, assume that the demon is totally selfless, solidly courageous and extraordinarily determined. These features make us believe the demon as a role-model on many different aspects, and thus the features give us a consideration for admiring the demon. Based on Hieronymi's understanding of reasons, now we have resources for differentiating between distinct kinds of considerations.

1.3.6 Plausibility of the Accounts of Right Kinds of Reasons

After clarifying the three theories of reasons of the right kind, I need to explain why these candidates are equally plausible. Here is my argument:

(1) If a theory of reasons of the right kind can differentiate between emotional appropriateness and prudential appropriateness, then the theory fulfills one of the necessary conditions for being a plausible theory of reasons of the right kind.

(2) Hieronymi, Schroeder, and Lang and Samuelsson's theories can distinguish between these different kinds of appropriateness.

Thus,

(3) Hieronymi, Schroeder, and Lang and Samuelsson's accounts fulfill one of the necessary conditions for being a plausible theory of reasons of the right kind.

Premise (1) is needed because of the conflation problem. This issue is that one type of appropriateness can be conflated other types of appropriateness (Sauer, 2011, 112).

The following example illustrates how emotional appropriateness can be conflated with moral appropriateness:

> John the Great is an ambitious dedicatory king who has a singular, personal and unshakable determination, namely, conquering the world with his massive army formed by huge groups of innocent civilians.

It is appropriate to admire his singular unshakable determination because it is admirable (or has properties that constitute the property of being admirable). But it is appropriate not to admire his determination because such a determination is selfish as he decides to achieve his personal goal by sacrificing many innocent people.

D'Arms and Jacobson also illustrate how emotional appropriateness can be conflated with prudential appropriateness:

> "Imagine that you have a rich and generous but touchy friend, who is extremely sensitive about his friends' attitude toward his wealth. If he suspects you of envying his possessions, he will curtail his largesse. That is a good reason not to envy him, […] but surely it doesn't speak to whether his possessions are enviable. (D'Arms and Jacobson, 2000, 731)

Neo-sentimentalism holds that to judge an object has a particular value is to judge that it is appropriate to have a certain emotion towards the object, but neo-sentimentalism needs extra resources to differentiate between distinct kinds of appropriateness, and theories of right kinds of reasons play a role in identifying distinct types of appropriateness. Furthermore, Deonna and Teroni (2012, 7) suggest that it is needed to distinguish between emotional appropriateness and other types of appropriateness such as prudential appropriateness.

Lang and Samuelsson's theory can distinguish between emotional appropriateness and prudential appropriateness. Their theory holds that your friend's wealth is enviable if and only if the wealth has properties other than its property of being enviable that give us a reason to envy the wealth, as long as the properties would still give us a reason to envy the wealth in the absence of any consequences of envying the wealth. Taking benefits from your rich friend is a consequence of not envying the wealth, thus, this talking is not a right reason for not envying the wealth.

Schroeder's theory can distinguish between emotional appropriateness and prudential appropriateness because of the criterion "the reason is shared necessarily by anyone who engages in A" in (SRKR). Taking benefits from your rich friend is not a right reason for not envying the wealth since this reason is not shared necessarily by anyone who engages in not envying the wealth. Some people are just not interested in taking benefits from your friend.

Hieronymi's account can distinguish between emotional appropriateness and prudential appropriateness. If your friend's wealth is enviable (or has properties that constitute the property of being enviable), then you have a reason for envying the wealth, and this reason bears on the question whether your friend's property is enviable (or has properties that constitute the property of being enviable). If your friend would give you fewer benefits if you envy your friend's wealth, then you have a reason for not envying the wealth, and this reason is a consideration which bears on the question whether it would be beneficial to envy your friend's wealth. The first consideration bears on an emotional question, whereas the second consideration bears on a prudential question, and thus we can distinguish between emotional appropriateness and prudential appropriateness in this case.

It is worth saying that theories of right kinds of reasons can make a distinction between emotional appropriates, prudential appropriateness, but neo-sentimentalists do not have to differentiate between emotional appropriateness and moral appropriateness here. If we endorse neo-sentimentalism, then to judge that it is immoral to take an action is to judge that it is appropriate to have a

negative emotion towards the behavior. In other words, the neo-sentimentalist reduces moral appropriateness to emotional appropriateness.

1.4 Objections towards Neo-Sentimentalism

Although the notion of appropriateness helps neo-sentimentalism avoids the problem of the recalcitrance of emotions and judgments, the theory apparently faces at least two objections because introducing the concept of appropriateness seems to entails that emotions are no longer important in this picture. It is apparently the case that reason alone is sufficient to provide one moral knowledge and motivate one to act morally. To defend the importance of emotions in the framework of neo-sentimentalism, I shall articulate and reply to two possible objections. Emotions have three different kinds of importance. The first kind is that emotions are a more cognitive-economical way to make moral evaluations. The second kind is that emotions make us sensitive to the values of the objects in environments. The third kind is that emotions motivate us to act morally. I shall articulate the first kind of importance in section 1.4, and clarify the other kinds of importance section 2.

1.4.1 An Epistemic Objection and Reply

If we accept neo-sentimentalism, then to judge that an action is morally wrong is to judge that it is appropriate to have a negative emotion towards the behavior, without actually having an emotion. To be appropriate to have a negative emotion is to have a right reason (in either Lang's, Schroeder's, or Hieromyni's senses) for having the emotion. A natural objection is that the right reason for having the emotion would also say what non-moral properties would constitute

moral properties. In other words, it is another way which bypasses emotions to gain moral knowledge. It leads us to believe that it is epistemically superfluous to gain moral knowledge via emotions.

I do not deny that reason can bypass emotion to make moral judgments, but I attempt to claim that emotions are a more efficient way to make moral evaluations. Such moral evaluations are implicit, which means that evaluations are manifested via something other than judgments, such as consciously felt physiological reactions towards objects. (I shall explain the meaning of evaluations as consciously felt physiological reactions in section 2). In psychology, the dual-process theory, a view which holds that mind is divided into two systems is widely accepted. System 1 involves automatic heuristic-based responses, and is fast, cognitively economical, and relatively narrow in scope, whereas system 2 involves conscious deliberation, and is slow, cognitively resource-demanding, but more flexible (Holton, 2009, 54). Spripada (2010, 795) suggests that system 1 includes "simply associative operations, and has access to limited information, whereas system 2 includes linguistic or logical representations, and has access to larger and more global stores of information." And Levy (2010, 145) claims that system 1 is undemanding of cognitive resources whereas system 2 is demanding of them. The fewer cognitive resources one has, the more difficult it is for one to function under system 2; the resources are depleted only if one is unavoidably switched to system 2. It is suggested that cognitive resources are related to glucose since the operation of the brain requires a tremendous amount of blood sugar, and the processes of the functions of system 2 burn glucose quickly (American Psychological Association, 2012).

Also, Greene (2007, 322) developed a dual-process theory of moral judgments. The dual process theory holds that there are two different ways to do moral reasoning in the brain. The first way is to make automatic and imminent reactions to actions that violate morality. These reactions are implicit and one might not consciously get access to the causes of the reactions. Greene believes that the automatic and imminent reactions are activated by the emotional areas in the brain. The second way is to do conscious and deliberative moral reasoning. Such reasoning is different from the emotional responses to the violation of morality. The reasoning is instead utilitarian: it focuses on how to maximize the benefits and minimalize the losses in the overall picture. Furthermore, Greene states that people who suffer from the damage in the ventromedial prefrontal cortex (VMPFC), a part which is related to the activation of emotions in the brain, tend to make utilitarian reasoning and decisions. Greene's theory matches the descriptions of system 1 and system 2 given by the philosophers mentioned above, saying that human beings have two ways to make moral evaluations.

If I attempt to introduce Greene's theory to my picture, then I need to solve a puzzle: If psychopaths do not concern the well-being of others because of the lack of moral emotions like empathy, then why psychopaths would make utilitarian moral decisions, which would not only benefit their own interests but also others'? Koenigs et al. (2011) studied in a comparison between psychopaths and non-psychopaths. They further divided the group of psychopaths into the group of psychopaths with a high level of anxiety and the group of psychopaths with a low level of anxiety. They invited the subjects to make moral judgments in hypothetical scenarios with moral dilemmas. The moral dilemmas describe "personal" and

“impersonal” actions. By “personal” it means actions that directly involve physical harms (such as “pushing one person off a bridge to stop a runaway train car from hitting five people”), whereas by “impersonal” it refers to actions that indirectly involve harms or cause harm remotely (like “pulling a switch to divert a runaway train car from hitting five people”). Both groups of psychopaths preferred taking impersonal actions. Yet only psychopaths with a low level of anxiety were in favor of executing personal harms if the harm would maximize the welfare of the overall individual.

To answer this question, it seems that Blair's speculation provides us a clue. Blair (1997) claims that psychopaths are unable to distinguish moral rules from conventional rules, yet they tend to treats both kinds of rules as inviolable as moral rules. Blair speculates that psychopaths attempt to persuade people that they are morally healthy. This founding suggests that psychopaths are dishonest and even cunning. Based on the deviousness of psychopaths, I am tempted to say that they express utilitarian moral reasoning and decisions because they realize that making and expressing utilitarian moral decisions can maximize their own interests. Instead, if they make and express decisions that only concern their own interests explicitly, then people would avoid having interactions with them, and thus they cannot maximize their own benefits because being unwelcome in society. After briefly presenting the dual-process theories and ego depletion, I argue as follows:

(4) Making moral evaluations via emotions is a process under system 1.

(5) Making moral evaluations via reason is a process under system 2.

(6) Processes under system 1 are more cognitively economical than processes understand system 2.

Thus,

(7) Making moral evaluations via emotions is more cognitively economical than making moral evaluations via reason. (From (4) to (6).)

Premise (4) is supported by evidence about decision-making and emotions. Bechara et al. (2000, 297) invented "the gambling game" and invited subjects to play this game to investigate the emotional influences on decision-making. The game simulates real-life situations which include uncertainty, rewards, and punishments, and the goal of the game is to get benefits based on the loaned money. The game had 100 rounds (the subjects were not told about this) and there were four tables named as A, B, C, and D. In each round they were allowed to select only one card from only one of the tables and were allowed to move to another table at any moment as they wanted. The game was followed by a series of rewards and punishments which were only known to the examiners. If subjects took a card from table A or B, they would gain $100 whereas if they took a card from table C or D, they would only gain $50. Nevertheless, the punishments which were unknown to the subjects were meted out more on table A and B than table C and D. Every 10 rounds, the subjects would lose $1250 on tables A and B, where they would lose only $250 on tables C and D. Measuring in long-term, table A and B were more disadvantaged because the subjects would lose $250 from table A and B in every 10 rounds whereas they would gain $250 on table C and D.

After the tenth round, the subjects were asked whether they had an idea of how the game operated. Meanwhile, their skin conductance responses (SCRs), which indicate emotional responses, were measured. In the first ten rounds, the subjects had mainly chosen the cards on table A and B, and thus they suffered their first losses. After the fiftieth round, the subjects started to choose cards from table C and D more frequently, and their anticipatory SCRs were activated stronger right at the moment before choosing card from table A and B than table C and D. Interestingly, at this stage the subjects reported that they had no idea of what was happening. After the eightieth round, Bechara, and the others claim that the subject has a "hunch" about what was happening and began to mainly choose a card from table C and D. The subjects reported that they *liked* tables C and D and *guessed* that the tables were more rewarding than tables A and B *without certainty*.

Here is another story which is similar to the card game. A firefighter *felt* that the situation was *not right without knowing what was going on exactly*:

> It is a simple house fire in a one-storey house in a residential neighborhood. The fire is in the back, in the kitchen area. The lieutenant leads his hose crew into the building, to the back, to spray water on the fire, but the fire just roars back at them. ''Odd,'' he thinks. The water should have more of an impact. They try dousing it again, and get the same results. They retreat a few steps to regroup. Then the lieutenant starts to feel as if something is not right. He doesn't have any clues; he just doesn't feel right about being in that house, so he orders his men out of the building—a perfectly standard building with nothing out of the ordinary. As soon as his men leave the building, the floor where they had been standing collapses. Had they still been inside, they would have been plunged into the fire below. (Klein, 1998, 32)

In the card game, it seems that the subject did not deliberate because they reported that they did not know what was happening in detail, and they claimed that they *like* tables C and D and that they *guess* tables C and D were safer.

Moreover, they naturally changed their behavior, moving to tables C and D from tables A and B. In the firefighting case, the lieutenant made the decision to leave the building based on his emotional response which indicated the heat of the fire and the lack of noise which was relative to the fire, but the lieutenant had not been aware of what spurred his feeling until Klein and his colleagues gave him their analysis (Holton, 2009, 63). These likings, guessings, and feelings fit the descriptions of system 1, and thus we have a reason to believe that these and similar emotional processes are governed by that sort of cognitive system.

Premise (5) is obviously true since it fits our daily experience. If we have finished our dinner and are about to pay the bill equally, dividing the total amount of the bill requires deliberative mathematical processes, no matter how simple the calculation is. Also, we have a good reason to believe that logical operations are under system 2 because rational calculation and making choices are ego depleting (Levy, 2010, 145).

Premise (6) is clarified by the notion of ego depletion. If making moral evaluations by emotions is a process under system 1, then making moral evaluations in this way does not require many cognitive resources. Instead, if one makes moral evaluations under system 2, then one makes moral evaluations in a cognitively resources-demanding way. Practically speaking, if both ways are channels to moral evaluations, and making moral evaluations is more economical under system 1, then it is unclear why making moral evaluations via emotion is not a preferred choice, although we can make moral evaluations by reason as well.

It needs to be mentioned that emotions and reason are equally reliable in terms of generating moral evaluations. Reliabilism for justification holds that a belief is justified if and only if the belief is produced by a reliable belief-making process (Becker, n.d., introduction). The examples of the reliable belief-making process include perception, solid memory, calculating, whereas guessing, wishful thinking are the examples of unreliable processes. The former kind of belief-making processes is reliable because they tend to produce true belief whereas the latter kind tends to produce false belief. We can also apply reliabilism to emotions. if an emotion is produced by a reliable emotion-making process, then the emotion-generated moral evaluations are equally justified as the judgments deriving from the operation of system 2 since reliably generated emotions are likely to be correct and thus the correspondent moral judgments are likely to be true.

It needs to explain why emotion-generated evaluations and the judgment driving from the operation of system 2 are equally justified. Both emotions and value judgments are mind-to-world, which means that they represent something outside the external world. To allow one's emotions and judgments to represent something, there must be reliable emotion/judgment-making processes as proper channels between one and the external world which are shared by both kinds of mental states. If I see John and hear the sounds of his footsteps, then my belief that John is there is justified since the belief is formed via my visual and audio perceptions that are proper connections to the external world. If I see the big claws and sharp teeth of the dog and hear that the dog is barking at me, then my fear of the dog is justified as well because the emotion is formed via the proper channels.

In section 2 I shall propose a more sophisticated version of reliabilism called *agent reliabilism*.

1.4.2 A Motivational Objection and Reply

Suppose that some workers are unavoidably working right outside your office in the department of philosophy. You are annoyed by the noise they make, and it seems that you have a right reason to feel annoyed because the properties of the noise would constitute the property of being annoying. However, at this moment, you suppress your emotion because you judge that the workers are just doing their job, and you ought not to show your ire towards people who are just doing their jobs unavoidably outside your office. This "ought" seems to give you some motivating force not to unleash your ire, which means that there is another source other than emotions, perhaps reason, to motivate you.

Although I disagree with that Prinz's understanding of the notion of appropriateness, we can explain the statement that there is another source other than emotions to provide motivation. The "ought" in the scenario can be understood as a second-order emotion. It is true that you feel annoyed by the noise brought by the worker, but you are motivated not to unleash the feeling of being annoyed because you have another negative emotion towards your emotion of being annoyed. Such second emotion can explain away the appearance of another source of motivation. Please note that I do not adopt Prinz's understanding entirely because the appropriateness of emotions should be analyzed in terms of reasons of the right kind.

Even if I cannot deny that there is a source other than emotions which provide us moral motivation to act, the motivational importance of emotion can still be preserved in a weaker sense. If one suffers from ego depletion, then one's reason cannot function effectively and efficiently, which means it is difficult for one's reason to deliberate and motivate one to act morally. But if one encounters ego depletion, one will be switched to system 1 even one is not willing to. Emotional processes are processes under system 1, which means that one can still be morally motivated by emotion in the condition of ego depletion. Moreover, system 1 is activated most of the time in one's daily life, whereas one only engages in system 2 and thus are motivated by reason occasionally. Emotions play the motivational role mainly in time.

2 The Bodily Attitudinal Theory of Emotions

In order to preserve the importance of emotions after introducing neo-sentimentalism, I suggest that there are three kinds of importance. The first kind is that emotions are a more cognitive-economical way to make moral evaluations. The second kind is that emotions make one sensitive to the values of the objects in environments. And the third kind is that emotions motivate one to act morally. I have articulated the first kind in section 1.4, and I shall explain the second and third kinds in this section.

In the previous section, I described neo-sentimentalism, a theory which holds that to judge that it is morally wrong to perform an action is to judge that it is appropriate to have a particular kind of negative emotion towards the action. These negative emotions include at least contempt, anger, and disgust.

According to Rozin et al. (1999, 576; Prinz and Nichols, 2010, 122), there are three different kinds of norms that correspond to these negative emotions, and which exist across many different cultures. These are community norms, autonomy norms, and divinity norms. Community norms are about public goods and social hierarchies. Autonomy norms prohibit harms against people. Divinity norms in non-secular societies forbid crimes against gods, whereas in secular societies forbid crimes against nature. The violation of community norms causes contempt, the violation of the autonomy norms brings anger, and the violation of the divinity norms spurs disgust. This model is called “the CAD model” (contempt, anger, disgust). I will use this model in my thesis.

I have analyzed moral judgments in terms of judgments of appropriate emotional responses, and the notion of appropriateness is further analyzed in terms of the right kinds of reasons for the attitude. What is still needed is a theory of the emotions. I will argue that Deonna and Teroni's bodily attitudinal theory of emotions provides us with an explanation of how moral judgments are motivating, and it can be used to give an explanation of how emotions and value judgments can be justified. Combining the bodily attitudinal theory with neo-sentimentalism, we can explain how moral judgments can be warranted.

2.1 The Central Thesis of the Bodily Attitudinal Theory of Emotions

The central thesis of the bodily attitudinal theory of emotions is as follows:

> Each emotion consists in a specific felt bodily stance towards objects or situations, which is correct or incorrect as a function of whether or not these objects and situations exemplify the relevant evaluative property. (Deonna and Teroni, 2012, 89)

While having an emotion, we experience a consciously felt physiological reaction towards an object. A consciously felt physiological response could make sense or not. It makes sense to have the physiological response if the object that the physiological reaction is directed at possesses the relevant value; otherwise, it does not make sense. According to Smith (2014, 98), the bodily attitudinal theory is composed of the following two claims: the first claim is that conscious physiological reactions are necessary emotions. The second claim is that to have an emotional attitude (a consciously felt physiological response) towards an object is to evaluate the object, in the sense that values determine whether it makes sense to have the

attitude towards the object. I shall explain each of these claims in more detail in the following sections.

2.1.1. Emotions as Felt Bodily Attitudes

Conscious physiological reactions *include* what the psychologist Nico Frijda calls *action readiness*, as well as changes in facial expressions and internal bodily systems. The following is how Deonna and Teroni describe conscious physical reactions:

> We should conceive of emotions as distinct types of bodily awareness, where the subject experiences her body holistically as taking an attitude towards a certain object. (Deonna and Teroni, 2012, 79)

Since the consciously felt physiological reaction consists of action readiness and the changes of facial expressions and internal bodily system, that is why a consciously felt physiological response is *holistic. Action readiness* is the conscious experience that one feels one's body as ready to take a particular kind of behavior. As Frijda (1988, 351) writes, "subjects report impulses to approach or avoid, desires to shout and sing or move, and the urge to retaliate, or on occasion, they report an absence of desires to do anything or a lack of interest, or feeling of loss of control." This observation provides the basis for the idea that action readiness at least partly constitutes emotions. While experiencing anger, one consciously feels one's body as ready to retaliate. While experiencing fear, one consciously experiences one's body as poised to escape from or eliminate the threat. While experiencing shame, one consciously feels one's body as prepared to hide oneself from other's eyes (Smith, 2014, 100).

Furthermore, Frijda (1988, 351) suggests that different kinds of action readiness can clearly differentiate distinct kinds of emotions. For instance, to have joy is to feel one's body as ready to cheer and approach an object; to be angry is to feel one's body as poised to remove or harm the object; to have shame is to have feel one's body as prepared to hide or cover ourselves.

Note that it is not the case that having an emotion is always having an interaction with an object directly. As Frijda writes:

> Joy, for instance, is a sense of pleasure plus the urge toward exuberance and contact-seeking. Anger is a sense of displeasure plus the urge to do some of the things that remove or harm its agent. Shame is a sense of displeasure plus the compelling desire to disappear from view. Sadness is a sense of displeasure plus the ebbing away of any urge except for the desire for the lost object or opportunity, which is known to be unfulfillable. Frijda (1988, 351)

Sometimes while experiencing an emotion, action readiness can be manifested in a way without having direct interaction with the object or event. For instance, while experiencing joy, I can celebrate *because* of an object, but my celebration can be manifested in a way without contacting the object directly, such as going to a bar for a drink with my friends.

It needs to be clarified that if one has a consciously felt physiological reaction, then one feels one's body as ready to do something, but not vice versa. We can feel our bodies as poised to harm or remove an object or engage in an event without anger, and we can feel our bodies as prepared to hide while playing hide-and-go-seek with my sister without shame. Action readiness is part of a physiological reaction but is not a sufficient condition for having a physiological response which are necessary to an emotion.

Different physiological responses may sometimes overlap, which means that the reaction may share the same kind of action readiness. For example, while experiencing intense anger and sadness, it is likely the case that one feels one's body as ready to get rid of the object that spurs one's emotions. But it does not mean that anger and sadness are the same emotion because the readiness to take a particular kind of action is merely a part of the physiological reaction. An emotion as a consciously felt physiological response is more than the readiness to perform a particular type of behavior. Although physiological response consists of action readiness, it also includes the changes of facial expressions (Frijda, 1988, 351). And as Deonna (2014, PowerPoint) mentions, a physiological reaction also consists of the changes in the autonomic nervous system and endocrine system. In other words, an emotion as a complete physiological response consists of action readiness and the changes of facial expressions, the autonomic nervous system, and endocrine system. Sadness and anger can be differentiated because they are distinct as a holistic consciously felt physiological reaction.

Some might object that we can experience emotion without action readiness, which means that action readiness is not a necessary condition for emotions. For example, one can fear a dog without being disposed to escape from or eliminate the dog. This objection can be explained away, however. If one has an emotion without exhibiting a particular kind of action readiness, the problem could lie in one's knowledge rather than action readiness. That is to say that one lacks the knowledge of how to respond to a particular kind of objects or events. Thus one cannot exhibit a particular kind of action readiness. As Frijda (1988, 351) mentions, it can be the case that our body is ready for action without knowing which action.

Instead, if one has the knowledge of how to respond to a particular kind of object or event, then while experiencing an emotion, one will possess the corresponding action readiness. Imagine a child who has never been taught how to deal with an aggressive dog. If unfortunately, the child is facing a dog, she cannot exhibit the corresponding action readiness because she has never been taught how to escape from or eliminate the dog. If she knows how to do so, she will have the disposition to escape from or even eradicate the dog as a threat.

Since the bodily attitudinal theory holds that emotions are consciously felt bodily attitude, the second kind of importance of emotions, namely, how to make us sensitive to the values of objects and thus make implicit value evaluations. Although one might not be able to access to why properties other than values give us a reason to have emotions towards objects, given the environments in question are so complex, consciously felt bodily attitudes make one aware of the values and thus make implicit evaluations. The bodily attitudes make us aware of them simply because the attitudes are consciously felt. And such implicit evaluations should be understood as behavioral reactions towards the objects. Looking back to the firefighter story, he was not able to access to the properties other than dangerousness that give him a reason to fear the building since the environment was so chaotic and dramatic. Yet he had felt that something was wrong without knowing why, and thus commanded his subordinates to leave the building. What he firefighter did can be explained by that he had a consciously felt attitude towards the building, and thus aware of the dangerousness of the building, then he made an implicit evaluation, namely, leaving the building.

The notion of action readiness is crucial to explain the third kind of importance of emotions, namely, how emotions motivate one to act morally. Emotions are motivating because they contain particular kinds of action readiness. If one judges that it is morally wrong to take an action, then one appropriately has a negative emotion towards the behavior, and with actually having the negative emotions, one is also motivated to act morally. A negative emotion includes adverse action readiness. By "negative" I mean stopping and preventing any further interaction or happening related to the object. We usually see anger and sadness as negative emotions. If one experiences anger, then one feels one's body as ready to retaliate, and retaliation prevents the further provocation given by the related object. If one experiences sadness, then one feels one's body as poised to evade the object, and evading stop one's interaction with the happening. Putting all the things together, if one judges that it is morally wrong to take an action, then one appropriately has a negative emotion towards the behavior, and with actually having the negative emotion, which means one has the behavioral disposition to cease and prevent any further interaction or happening related to the behavior.

In summary, the bodily attitudinal theory holds that emotions are consciously felt physiological reactions that include action readiness and the changes of facial expressions and internal bodily systems. Being ready to take a particular kind of behavior does not mean always interacting with an object directly. Although having an emotion-constituting physiological reaction entails being prepared to perform a particular kind of behavior, being poised to carry out some action is insufficient for having a consciously felt physiological response. Some emotions share the same readiness to take a particular kind of behavior, but they

are different because they have a distinct combination of the changes of facial expressions and internal bodily systems. While having a conscious physiological reaction, one's body may exhibit no action readiness since one does not know what action to take. If one's body is ready to take some action and one knows what to do, then action readiness is exhibited.

2.1.2. Attitudes and Evaluation

Having clarified what it means for emotions to be consciously felt bodily attitudes, I shall explain the second claim. The relation between an attitude and its content needs to be introduced and explained because the relation is crucial to illustrate why we should favor the bodily attitudinal theory over rivals such as the judgmental and the perceptual theories of emotions.

We can start by considering familiar propositional attitudes. These attitudes are helpful to illustrate the difference between an attitude and its content. Examples of propositional attitudes include *believing that snow is white*, *hoping that Mt Rosea is twelve miles high*, and so forth (Oppy, n.d., introduction). The attitudes, such as *believing* and *hoping*, are different from their contents *snow is white* and *Mt Rosea is twelve miles high.* Attitudes are evaluations of the content, which means that to have an attitude towards the content is to evaluate the content in a certain way. In the case of belief, when I believe that snow is white, I have an attitude *believing* (considering as true) towards the content *snow is white*, which means that I evaluate the content as true. Similarly, in the case of emotion, when I fear the dog, I have an attitude of *fear* (considering as dangerous) towards the content *the dog*, which means that I evaluate the dog as dangerous.

We need to clarify, in what sense beliefs and emotions are evaluations towards their contents. In one mistaken understanding, beliefs are evaluations because truth is a part of the content. In other words, when I believe that *snow is white*, I have the content that *it is true that snow is white*. This sense is mistaken because the view entails that no creatures would be able to form beliefs without the concept of truth. Call this the content view. Instead, beliefs are evaluations in the sense that truth is the *formal object* of belief, a standard that decides whether the attitude of believing is correct, but is not a part of the content. The attitude *believing* towards the content *snow is white* is correct if and only if snow is white. Call this the formal object view.

Similarly, one could hold that emotions are evaluations because values are a part of the content. When I fear the dog, I have an emotional attitude towards the dog that is an evaluation because dangerousness is a part of the content *the dog.* Instead, on the formal object view, emotions are evaluations because values are the formal objects of these emotions, standards that determine whether the attitudes are correct. The attitude of fear towards the content *the dog* is correct if and only if the dog is dangerous, but dangerousness is not part of the content of the fear.

As the difference between the content view and the formal object view has been clarified, we need to choose one. I shall explain why we should be in favor of the formal object view but not the content view.

According to the content view, to have emotional attitudes towards contents is to evaluate the contents in the sense that values are a part of the content. One version of this view holds that we *judge* that the contents have certain

values. The judgmental theory holds that value judgments are necessary for having emotions, which means that the value judgment is a necessary part of the emotion (Deigh, 2010, 25). For instance, the judgment that the dog is dangerous is necessary having the fear of the dog. The judgment that the man is offensive is necessary for having the anger at the man. An emotion cannot exist without the corresponding value judgment. Fear of the dog cannot exist without the judgment of dangerousness, and anger at the man cannot exist without the judgment of offensiveness. Nonetheless, it does not mean that having a value judgment entails having an emotion since it is possible to have a value judgment without an emotion unless the emotion and the value judgment are identical.

A major objection to the judgmental theory is the problem of the recalcitrant emotions. Sometimes one has emotions without the corresponding value judgments; sometimes one has emotions that are contrary to one's value judgments. It follows that value judgments are neither necessary nor sufficient for having emotional responses (D'Arms and Jacobson, 2003, 129). To illustrate, it is possible to fear a dog even while one judges that the dog is not dangerous, and one can fail to fear what one judges to be dangerous.

A simple way to defend the judgmental theory is to say that the fearful person judges that the dog is dangerous but at the same time judges that the dog is not dangerous. In other words, she holds contradictory judgments, and thus experiences a recalcitrant emotion. Nonetheless, it is implausible to the claim that we have a rebellious emotion because of the conflicting judgments. The defenders of the judgmental theory must provide evidence for the claim that we have

conflicting judgments while experiencing a rebellious emotion. Otherwise, it is merely an ad hoc hypothesis. It is unconvincing to accept ad hoc explanations since they weaken the testability of theories, which means that ad hoc hypotheses make theories unable to be confirmed or falsified.[1] Another further objection towards the straight reply is that we usually experience recalcitrant emotions whereas we seldom have contradictory judgments, which means that both kinds of states are asymmetric in terms of frequency. The judgmental theory, therefore, remains unappealing.

Some scholars, such as Greenspan, offer a modification of the judgmental theory, suggesting that emotions are "perceptions" of values (D'Arms and Jacobson, 2003, 130). To have an emotional attitude is to *perceive* the content as having a particular value. To have the attitude fear towards the content the dog is to *perceive* that the dog is dangerous. This view is called the perceptual theory. Please note that it does not mean that when one has an emotional attitude one *literally perceives* the evaluative content. It should be understood as a quasi-perceptual state. Quasi-perceptions are similar to visual and audio perceptions like looking at a red cube and listening to a banging sound, because both quasi-perceptions and perceptions share some similarities such as incorrigibility and having vivid phenomenology (Baker, 2014, 1). Since emotions as quasi-perceptions are relatively incorrigible, the recalcitrance of emotions is explained. Imagine that one perceives a big pink rabbit in front of the door, and everyone says that there is in fact no such

[1] Suppose an astrophysicist's theory predicts that there is a planet running around the sun in our solar system. Yet, via our finest and most sophisticated telescopes, we observe no planet there. The astrophysicist then adds an ad hoc hypothesis to protect her theory: the planet is invisible, so we cannot see it via our finest telescopes. The ad hoc hypothesis makes the theory less testable even not testable at all. if visual observation were the only possible test.

rabbit at all. Then one has a reason for believing that one's perception of the big pink rabbit is an illusion, but the illusion can remain even if one makes a judgment that is contradictory to the illusion.

Still, it is implausible to accept the perceptual theory. According to Baker (2014, 5) and Schroeder (2008, 122), perceptual theories of *desire* roughly are the views that to desire something is to quasi-perceive it as good. This kind of accounts face an objection: they are incompatible with the fact that infants and animals have desires because quasi-perceiving something as good requires the concept of *good*. Applying this objection to the perceptual theory of emotions, the concepts of values are needed to have quasi-perceptions of values. Perhaps some argue that infants and animals do not need the concepts of values to have perceptions of evaluative properties. Although it is acceptable to say that children and animals have simple concepts like the concepts of a red cube and a pink rabbit, it seems improper to say that they have the concepts of value since they are more complicated than those simple concepts. The concepts of values are more complicated in a way that they are too abstract for them to handle. The concept of a red cube and a pink rabbit is not abstract because infants and animals can know what they are by seeing a red cube and a pink rabbit. But the concepts of values such as wrongness are too abstract to them. To know what wrongness is, they need to know what harm and deception and other actions that we consider as wrong are.

This problem is welcome news for my project: the perceptual theory is not a good partner to neo-sentimentalism because together they face a circularity problem. As D'Arms and Jacobson (2003, 127) write, "…the sentiments adduced to

explicate moral concepts already involve the very content they are supposed to explain." Neo-sentimentalists attempt to analyze the moral concepts in moral judgments in terms of appropriate emotional responses, and the perceptual theorists want to explain emotions in terms of quasi-perceptions of evaluative properties. But now we need to introduce the concepts of values to have the quasi-perceptions of values, which mean that the perceptual theorists use something that neo-sentimentalists attempt to explain to explain what emotions are.

In summary, according to the content view, to have emotional attitudes towards the content is to evaluate the content in the sense that values are a part of the content. In other words, emotions consist of value judgments or quasi-perceivings. The judgmental theory should be rejected because of the problem of recalcitrant emotions. The perceptual theory fares better on this score, but it is incompatible with the fact that infants and animals have emotions. This is welcome news for sentimentalists about moral judgment, as sentimentalism cannot be combined with either of these theories without circularity.

Does the bodily attitudinal theory face the problem of circularity as well? Some might think that the bodily attitudinal theory faces the same problem because concepts of values are needed to have consciously felt physiological reactions. I do not believe so. The bodily attitudinal theorists analyze emotional attitudes in terms of conscious physiological responses, and having these reactions requires no concepts of values; thus, the theory does not exclude the fact that infants and animals have emotions. Also, the bodily attitudinal theory does not result in circularity when combined with neo-sentimentalism, because the concepts

of values are analyzed in terms of appropriate emotional responses, and the emotional responses are further analyzed as instances of physiological reactions.

2.2 Justifying Emotions and Value Judgments

This is Deonna and Teroni's account of when emotions are justified:

> An emotion is justified if, and only if, in the situation in which the subject finds herself, the properties she is (or seems to be) aware of and on which her emotion is based constitute (or would constitute) an exemplification of the evaluative property that features in the correctness conditions of the emotions she undergoes.
>
> (Deonna and Teroni, 2012, 97)

My fear of a dog is justified if and only if my fear is (or seems to be) based on my awareness that the dog has the relevant non-evaluative properties, such its sharp teeth, claws, and impulsive behavior that constitute (or would constitute) dangerousness. Introducing the phrase "would constitute" is required because whether the non-evaluative properties constituted the values depends on who one is and what the situation is. If one is Ironman, then the non-evaluative properties probably do not constitute dangerousness since one is wearing a super armor. However, if one is just an ordinary person without protection, then the non-evaluative properties probably constitute dangerousness. The word "would" merely means that it is not the case that the non-evaluative properties constitute the evaluative properties to all subjects in all circumstances. Seeming roughly means that we think something in the way it is default. I shall explain what seeming means with phenomena conservativism below.

Also, Deonna and Teroni believe that a justified emotion is sufficient to justify the corresponding value judgment. If one is (or seems to be) aware of the

non-evaluative features possessed by the particular object which constitutes (or would constitute) the relevant evaluative property, then one's emotion is justified. With the justified emotion, if one has no reason to reject the seeming, then one is also prima facie justified in believing the corresponding value judgment. If one is or seems to be aware of that the non-evaluative properties of the dog which would constitute dangerousness, then one is justified to have a fear of the dog. And if one has no reason to refuse the seeming, then one is prima facie justified to believe that the dog is dangerous. The phrase "prima facie" is needed because a seeming is not a decisive justification for emotion because the seeming does not exclude other possibilities.

Huemer's (2005, 99) phenomenal conservativism can be used to help illustrate Deonna and Teroni's account. According to Huemer (n.d., section, 1a), if it seems to subject S that P and there is no reason or evidence against P, then S is to some extent justified in believing P. The meaning of the phrase "it seems to subject S that P" should be understood broadly. It can mean perceptual, intellectual, memorial, introspective appearances. If it appears to be that the sky is blue, and there is no reason or evidence against this seeming, then I am justified in believing that the sky is blue. Furthermore, an appearance is defeated if the appearance does not match reality since the appearance could be hallucinations or false memories and so forth. As Huemer writes:

> Appearances sometimes fail to correspond to reality, as in the case of illusions, hallucinations, false memories, and mistaken intuitions. Most philosophers agree that logically, this *could* happen across the board – that is, the world as a whole could be radically different from the way it appears. These observations do not conflict with phenomenal conservatism. Phenomenal conservatives do not hold that appearances are an infallible

> source of information, or even that they are guaranteed to be generally reliable. Phenomenal conservatives simply hold that to assume things are the way they appear is a rational default position, which one should maintain unless and until grounds for doubt ("defeaters") appear. (Huemer, n.d., section 1b)

Applying phenomenal conservativism to Deonna and Teroni's account, if it seems to one that the non-evaluative properties that would constitute the value is present, and one has no reason for refusing the seeming, then one is prima facie justified in believing the corresponding value judgment. The idea is that the justified emotion entails that it seems that a particular object possesses the non-evaluative property that would constitute the value. In other words, a justified emotion is a defeasible justification for the value judgment. If an emotion is unjustified, then we have no reason to believe that the particular object possesses the non-evaluative properties that would constitute the value, then it is also unjustified to believe the value judgment.

2.3 Bodily Attitudes and the Teleological Theory of Representation

When experiencing an emotion, one feels the changes of one's facial expression, muscular movements, and the internal bodily systems. For example, while experiencing anger, one feels that one's face and muscles are in tension, the increase in one's heart rate and breathing and so forth. The psychologist William James believes that these bodily phenomena are emotions. Moreover, these bodily phenomena are not the results of emotions but emotions themselves. As James writes:

> My thesis, on the contrary, is that the bodily changes follow directly the perception of the exciting fact and that our feeling of the same changes as

they occur is the emotion. James (1884, 189)

The meaning of perception here is different from the meaning of quasi-perception in the perceptual theory of emotions mentioned previously. In James's theory, perceptions literally mean bodily sensations. To support this theory, James introduces a thought-experiment:

> I now proceed to urge the vital point of my whole theory, which is this. If we fancy some strong emotion, and then try to abstract from our consciousness of it all the feelings of its characteristic bodily symptoms, we find we have nothing left behind, no "mind-stuff" out of which the emotion can be constituted, and that a cold and neutral state of intellectual perception is all that remains. (James, 1884, 189)

James's theory has one selling point, that is, the theory is not cognitively demanding and thus, unlike the judgmental and perceptual approaches, it does not exclude the fact that infants and animals have emotions (Deonna and Teroni, 2012, 64). However, the theory has a serious flaw which is that it fails to explain the intentionality of emotions. Understanding emotions only as the changes of the bodily perceptions is incompatible with our daily speaking of emotions, such as "I am angry *at* John" and "She is sad *because* the cat is dead." Emotions must be about something, and the theory does not mention the link between emotions and the objects that they are directed at.

The bodily attitudinal theory, however, preserves the intentionality of emotions, as its central thesis is that every emotion consists of a consciously felt physiological reaction that is *directed at* an object or event. But what makes the conscious physiological responses directed at something? We need a theory of

intentionality to illustrate the link between emotions and the objects that they are directed at in the bodily attitudinal theory.

We believe that our mental states represent something in the world. But why are they about something? What is the representational relation between the mental states and the world (Peters, 2014, 273; Millikan, 2003, 3)? Applying the question to the intentionality of bodily attitudes, why are they directed at contents or objects? What determines bodily attitudes directed at contents or objects? We need to introduce to a theory of representation as a complete theory of emotions is required to explain the intentionality of emotions. A plausible theory of representation is the teleological theory of mental content. This theory holds that a mental state represents something because of its biological function to indicate the object in question. I believe that this approach is suitable for explaining the intentionality of the emotions, since some believe that emotions are evolutionary products (Johnson, n.d., section 2), and the teleological theory also explains the representation of mental states in terms of evolution.

The teleological theory of mental content holds that a mental state in a creature represents an object if and only if the mental state is biologically selected to perform the function of detecting the object (Peters, 2014, 274). The concept of the *dog* represents dogs because the concept has the biologically selected function indicate dogs. It is just like that a thermometer represents the temperature because it has the designed function of detecting the temperature. The biological function of indicating dogs exists because the concept was selected to detect dogs and develop advantageous behavior for responding to dogs. Evolution shapes the selection over

a period, or by learning or conditioning during the lifetime of a creature. Applying this theory to bodily attitudes, such attitudes are directed at their contents or objects because they have their biological functions of indicating the content or objects, and evolution or learning shape the biological functions. Evolution develops our bodily attitudes so we are more likely to survive in hazardous environments and more social in cultivation. And to develop bodily attitudes by learning is helpful to achieve particular purposes. For instance, a boxer needs to learn to be angry at his opponent to win the competition.

The teleological theory is an “externalist” theory, which means the mental content is not determined by anything such as memory and seeming, in one’s mind consciously, but determined by what one encountered in the past. As Millikan writes:

> Thus naturalistic teleological theories are "externalist" theories of mental content. They imply that the content of one's thought is not determined by anything before one's mind or within one's consciousness or even within one's head. Just as actually remembering something, rather than merely seeming to remember it, does not happen wholly within one's present head but requires that one has previously encountered that thing, thoughts that are about something actual also require the right sort of history. It would be possible for a teleologist to avoid this externalism only with a non-historical and also non-environment-relative account of the nature of biological functions. (Millikan, 2003, 5)

A major objection towards the teleological theory of mental content is the swamp-man objection. If the teleological theory is true, then the representation of a mental state in a creature is determined by its biological function, which is selected for by evolution, learning or conditioning. However, we can imagine that the mental states in a creature represent something, yet their functions are not determined by evolutionary selection or learning (Peters, 2014, 275). Davidson (1987, 443)

mentions the following thought-experiment: suppose that Davidson is walking nearby a swamp and suddenly gets stuck by lighting. The lighting totally destroys Davidson, meanwhile coincidentally the materials from the swamp perfectly reassemble into a replica of Davidson at the molecular level. The replica of Davidson is behaviorally identical with Davidson, which means that the replica talks, walks, and writes like Davidson. Since the replica is made of the materials of the swamp, he is called "the swamp-man." We tend to think that the mental states of the swamp-man are representational. But the mental states were not selected for, nor are they the result of learning.

There are two responses to the swamp-man objection. Some teleological theorists deny that the swamp-man has representational states (Neander, 2006, 385; Millikan 1996, 7). Other teleological theorists deny that the swamp-man is insufficient to falsify the teleological theory because the swamp-man is not a creature in reality since the teleological theory is only about real creatures (Peters, 2014, 276).

To explain why bodily attitudes are directed at objects or contents, and why fear is directed at the dog, the teleological theory is a possible move. Bodily attitudes are directed at their objects or content because they have their biological function of detecting the objects and contents, and their selection history designs the function to develop appropriate evolutionary behavior for the species' flourishment. Perhaps, more specifically, emotions are biologically designed in order to promote pro-social behavior in civilized societies (Prinz, 2007, 80).

2.4 Objections towards the Bodily Attitudinal Theory of Emotions

2.4.1 Joel Smith's First Objection and Reply

Since the bodily attitudinal theory holds that consciously felt physiological reactions are necessary for having emotions, the proponents of the theory need to explain why there are some patients who have a rich emotional life, although lacking the abilities to feel bodily sensations and perform muscular movements (Smith, 2014, 101). A contemporary bodily theorist of emotions, Damasio (2006, 155), suggests that there is a device called "as-if loop" within the brain This device simulates the changes in the bodily feelings in cases which veridical bodily feelings are no longer available.

We can reasonably assume that the "as-if loop" device is Deonna and Teroni's default defense. Nevertheless, it seems that the "as-if loop" device cannot help defend the bodily attitudinal theory because Deonna and Teroni claim that emotions are essentially felt, yet Damasio explicitly refers to the three following states:

> A state of emotion, which can be triggered and executed unconsciously; a state of feeling which can be represented non-consciously; and a state of feeling made conscious, i.e. known to the organism having both emotions and feelings. (Damasio, 2000, 37)

Since Deonna and Teroni claim that emotions are essentially felt and the "as-if loop" device simulates bodily feelings unconsciously, it is unclear how they could use the device as a reply to this objection, and thus they accept both of the following statements, which creates a tension without a solution (Smith, 2014, 102):

> (a) Conscious emotion can persist in patients without any bodily feeling, veridical or otherwise.

(b) Emotional phenomena are bodily phenomena.

To reply to this objection, I will reject (a) and preserve (b). I believe that besides the "as-if" device phantom limbs sensation and phantom pain open another possibility for the bodily attitudinal theory.

Phantom limb sensations are common phenomena among a variety of individuals (Giummarra et al., 2007, 220). At least 98% of the patients experience phantom limbs after amputation, nerve damage, or spinal cord damage, and approximately one-fifth of children with inborn limb aplasia experience the phantoms as well. 80% of patients who had a limb removed reported that they experienced various kinds of phantom pain, including burning, tingling, throbbing and so forth. Some patients report that they have phantom sensations immediately after amputation, whereas some report that the sensations emerge years later. Phantom sensations may be the sensations of limbs such as arms, legs, breasts, and even internal organs. Phantom limbs are described as occupying space, which means that it feels like they have sizes, sharps, and postures, and patients claim that they can control the phantom limbs.

Phantom limbs sensations can be perceived normally or abnormally (Giummarra et al., 2007, 221). Some patients report that they can move their phantom limbs within a limited range whereas some claim that they can move the limbs to anatomically impossible positions. Patients who suffered from spinal cord damage tell that they perceive phantom limbs sensation in their lower body unrealistically, such as feeling "like the toes are all turned down under the bottom of the foot."

In the past, it was speculated that phantom limbs sensations are caused by the neuromas at the tip of stumps. However, this theory is insufficient to explain the sensations, since people who suffer from inherent limb deficiency sometimes feel phantom limbs as well. This suggests that there is a central representation of the limbs in the brain (Ramachandran and Hirstein, 1998, 1604).

Another theory, which explains the central representation, is called "the neuro-matrix theory" (Giummarra et al., 2007, 224). This theory holds that the attentions and the perceptions of the body are caused by configured activities based on the neuro-signatures inside of the brain. Neuro-signatures can be generated by a variety of causes, such as somatic inputs, visual inputs, and even the changes of the immune system. Although the neuro-matrix created by different inputs, the removal of limbs and the losses of other kinds of biological inputs can lead to a cortical reorganization, and the outburst of the activities of the neuro-matrix produce configured activities similar to the activity associated with pain and eventually cause conscious phantom pain. Whilst the neuro-matrix theory can explain that some phantom pain is caused by sources other than injury, lesion and the losses of limbs, the theory is difficult to test empirically (Giummarra et al., 2007, 224). The theory is unable to explain why phantom limb pain cannot be weakened by the disappearance and the spontaneous termination of phantom limb sensation, and why some patients do not feel phantom limb pain after amputation (Bittar et al., 2005, 401).

Although the fact that there is not yet a satisfactory theory which explains the nature of phantom limb pain and phantom limb sensations, these phantom

states are wide-spread phenomena. It is possible that patients with spinal damage experience "phantom emotions." By "phantom emotions" I mean that while those patients are reporting that they experience conscious emotions, they may be feeling the changes of the phantom sensations of their internal organs and the muscular movement that are longer physically linked to their brain. Given the majority of amputees and people who suffer from congenital limb deficiency experience phantom limb pain and phantom limb sensation, "phantom emotions" could well be a usual phenomenon. Thus, we can preserve the central thesis of (BAT) while explaining those patients with complete spinal cord damage. However, I have to make it clear that this is only a possibility because such speculation is not empirically verified and further investigation on the relations between phantom limb sensations and "phantom emotions" is needed.

2.4.2 Joel Smith's Second Objection and Reply

This second objection is about the correctness condition in the bodily attitudinal theory. Deonna and Teroni's account of the correctness condition is as follows:

> Fear of the dog is an experience of the dog as dangerous because it consists in feeling the body's readiness to act so as to diminish the dog's likely impact on it (flight, preemptive attack, etc.), and this felt attitude is correct if and only if the dog is dangerous. (Deonna and Teroni, 2012, 81)

While it is natural to say that it is correct to run if and only if the dog is dangerous, the meaning of correctness needs to be clarified further. What is meant by "correct"

in this context? Surely being correct cannot mean being true because action readiness cannot be true or false.

The following examples provided by Deonna and Teroni themselves vividly illustrate the seriousness of this problem: anger is the experience that one feels that one's body is ready to retaliate, and this anger is correct if and only if another person is offensive to one. But Jesus disagrees with such a saying as it is written in the Bible: if anyone strikes you on the right cheek, turn the other so (Matthew 5:39). It seems that you ought to agree with Jesus if you are a Christian.

Another way to clarify the meaning of correctness is to say that being correct means being rational, yet this can also be questioned: is it rational to execute retaliation if a person is offensive to you? In the spirit of G.E. Moore, if being correct is synonymous with being rational in this context, then we would not say no to the above question since it would be self-contradictory. Smith (2014, 104), as Deonna and Teroni (2012, 7) mention, suggests that the seriousness of this problem lies in differentiating emotional appropriateness, prudential appropriateness, and moral appropriateness.

I have already replied to this objection before in section 1. Since I propose that being correct means being appropriate to have an emotion, which means that one has a right reason to have the emotion, I shall not investigate this objection further. I wonder why this section is even necessary given that it goes nowhere and ends up depending on what was said in section 1.

2.4.3 Joel Smith's Third Objection and Reply

The third objection is that it is unclear how the awareness of the non-evaluative content in one's cognitive base justifies the corresponding evaluative attitude (Smith, 2014, 104). It is difficult to tell how the awareness of the impulsive behavior, the shape of the teeth, and claws possessed by a dog justifies the fear of the dog. Unless we have further information that a particular kind of non-evaluative properties is identified with a particular kind of evaluative property, it is implausible to think that the justification for emotions can travel via the content.

The straightest way of replying to this objection is to claim that a particular kind of non-evaluative property is identical with a particular kind of evaluative property, and thus, the non-evaluative content can justify the evaluative attitude. However, I shall argue that the identity relation cannot be proven empirically because the inference to the best explanation can prove the claim that water is H_2O whereas we cannot prove the claim that the non-evaluative properties are identical with the evaluative properties. The claim that water is H_2O can be proven by the inference to the best explanation because the claim possesses some explanatory power. Here is an observable fact:

> If we submit a sample of water to electrolysis, then hydrogen will be formed at the cathode and bubbles of oxygen will form at the anode. The mass of water decreases by the same amount that the mass of hydrogen and oxygen increase. Two moles of hydrogen are formed for every one mole of oxygen. (Huemer, 2005, 89)

Base on the observable fact, we can make the scientific hypothesis that water is H_2O because the hypothesis causally and counter-factually explains the observable fact. The causal explanation is that water is constituted by hydrogen and oxygen, and it is the only causal explanation to the fact. The counter-factual explanation is

that the fact would not happen if water were not identical with H_2O, and no other counter-factual explanation can explain the fact equally well.

However, the claim the non-evaluative properties are identical with evaluative properties cannot be proven by the inference to the best explanation. Assume that John kills many people, and we make the moral hypothesis that John is immoral. The moral hypothesis possesses no explanatory power since it cannot causally or counter-factually explain why John kills many people. The moral hypothesis cannot causally explain that because we can explain that John kills many people without using any moral terms (Huemer, 2005, 90). The moral hypothesis cannot counter-factually explain the fact because we can explain the fact equally well with a psychological story (Huemer, 2005, 90). For instance, a psychologist can say that if John's parents had not neglected him, then he would not have killed anyone.

Another less ambitious strategy is to claim that the particular kind of evaluative properties supervenes on the particular kind of non-evaluative properties, which means that if the non-evaluative properties change, then the evaluative properties change. Although the evaluative properties supervene on the non-evaluative properties, they are not identical with each other. The evaluative properties supervene on the non-evaluative properties because the non-evaluative properties exhaustively constitute the evaluative properties (Shafer-Laudan, 2003, 76). Imagine there is a lead statue. The piece of lead exhaustively constitutes the statute, yet they are not identical with each other since the piece of lead would remain even if the statute is destroyed. Since the non-evaluative properties

exhaustively constitute the evaluative properties, the non-evaluative properties can justify the evaluative attitudes.

Even if the supervenience strategy is also untenable, there is still a way of defending the justificatory account of the bodily attitudinal theory. It is unnecessary to justify the evaluative attitude via the awareness of the non-evaluative content, which means that one can bypass such awareness to justify the attitude. Deonna and Teroni seem to assume epistemic internalism in this picture, yet I think that epistemic externalism opens another possibility to justify the attitude. Briefly, epistemic internalism requires that one is able to cognitively get access to the justification for beliefs, whereas epistemic externalism holds that one does not need to cognitively get access to the justification, and allows the justification to be grounded outside one's cognition (Bonjour, 2002, 234; Poston, n.d., introduction).

An argument against epistemic internalism and for externalism is that unsophisticated epistemic subjects cannot have internalist justification (Bonjour, 2002, 242). These subjects include higher animals, young children, and cognitively unsophisticated adults. Higher animals can understand a wide variation of commands, to have desires, emotions, and even beliefs, it is unclear that they are sophisticated enough to have reasons for their beliefs. If an argument is markedly complicated, requiring one to divide the argument into many parts and know a lot of technical terms, then young children and some elderly people cannot cognitively understand the argument and thus have internalist justification. Furthermore, the chicken-sexers are experts who are capable of distinguishing between male and female chickens. It is reported they are unable to describe how to do the distinction,

yet they can sort chickens and believe that *this is a male and this is a female*. Since we tend to think that higher animals, young children, elder people and chicken-sexers have knowledge to a certain extent, it is either the case that knowledge does not require justification or that justification is not internal (Poston, n.d., section 3b). If there are cases that people have knowledge without internal justification, then analogously it seems plausible to believe that one can be justified to have the evaluative attitudes without internal awareness of the non-evaluative content in one's cognitive base.

One famous version of externalism is Goldman's Process reliabilism. This theory holds that a belief is justified if and only if the belief is formed by a reliable belief-making process (Goldman, 1979, 97). A belief-making process is a process that takes certain inputs such as visual experience and transforms them into certain outputs such as beliefs. A reliable belief-making process is a process that tends to produce true beliefs. Perceptions and memory are reliable belief-forming processes because they tend to produce true beliefs in most of the cases that they are applied, whereas guessing and wishful thinking are unreliable processes because they tend to produce false beliefs in most of the cases. By "tend" it means that justification is a matter of degree. The more reliable a belief-making process is, the more justified the belief produced by the process is. If I clearly remember that I ate noodles this morning, then my belief that I ate noodles this morning is more justified, whereas if I fuzzily remember that, then my belief is less justified.

Nevertheless, a reliable belief-forming process is neither sufficient nor necessary for justifying a belief (Greco, 2002, 292). Suppose John has a rare sort of

brain lesion which causes him to produce the belief that he has a brain lesion. However, he has no evidence for it, and even has evidence against it, because he has been given a medical analysis that claims that he is perfectly healthy by qualified neurologists. We tend to think that his brain lesion is a reliable way to form the belief, yet we do not think that he is justified to have the belief. This demonstrates a reliable belief-forming process is not sufficient for justifying beliefs. Assume that John is in a world full of massive deception created by a mighty demon. In this world, everything is exactly the same as in the actual world. John can form his beliefs by different processes like perceptions and memory. But we do not think that he forms these beliefs reliably because of the massive deception, yet we tend to think that he is justified to have the beliefs. Thus, a reliable belief-forming process is not necessary for justifying beliefs.

To avoid the brain lesion case and the mighty demon case, we can adapt Agent Reliabilism. This theory holds that a belief is B-justified (being justified for beliefs) if and only if the belief is a product of one's intellectual virtues as B-reliable processes (processes that tend to produce true beliefs). There are at least two understandings of intellectual virtues. Sosa understands intellectual virtues as cognitive abilities such as perception, memory, and introspection, whereas Montmarquet understands these virtues as personal traits such as intellectual courage and intellectual carefulness (Greco, 2002, 293). Agent reliabilism is able to explain the counter-examples to Goldman's reliabilism. In the brain lesion case, the belief caused by brain lesion is B-unjustified because it is not the outcome of one's intellectual virtues as B-reliable processes. In the mighty demon case, although one is being massively deceived, one is B-justified to have the beliefs because they are

the products of one's intellectual virtues as B-reliable processes. A belief is B-justified because the belief is the product of one's intellectual virtue as B-reliable processes. And one's intellectual virtue as a reliable belief-forming process likely produces true beliefs. Analogously, if an emotion is formed via one's intellectual virtue as E-reliable processes (processes that tend to produce correct emotions), then the emotion is E-justified (being justified for emotions) since the intellectual virtue as reliable emotion-forming processes tend to produce emotions which are likely correct. If we have no reason for questioning the E-reliability of the emotion-forming process, then the correspondent value judgment is B-justified as well.

A reason for justifying emotions in this way is that agent reliabilism can explain the accountability of having emotions. We can hold people accountable for having certain emotions (Prinz, 2007, 115). If one forms an emotion based on unreliable emotion-making processes, we say that one's emotion is not justified and one is accountable for having the emotion. Instead, if one forms an emotion based on one's cognitive abilities like perceptions, we say that one's emotion is E-justified and one is praiseworthy for having the emotion. In this context, emotion-making processes should be understood as cognitive abilities like perceptions but not personal characteristics like carefulness since for example, it seems that carefulness entails that one is making an emotion under the processes of system 2. In other words, one is making an emotion under some kind of intentional control, and thus my view is contradictory.

2.5 Moral emotions and Non-Moral Emotions

Suppose two boxers are fighting against each other in a boxing competition. Both of the boxers are angry at each other indeed, but they do not think that the other is acting morally wrong. From this example, we can tell there is a distinction between moral anger and non-moral anger, or in generally, there is a distinction between moral emotions and non-moral emotions. This distinction is important for justifying moral judgments as the boxing example tells that non-moral anger can be appropriate even if it would not make sense to judge the target of such anger to be morally wrong. The non-moral anger can justify the judgment that the person is offensive. But being offensive is not always being morally wrong. Instead, we can legitimately and naturally claim that the moral anger justifies the judgment that the person is morally condemnable. To justify the judgment of wrongness, the anger must be a moral anger. What makes an anger a moral anger?

The answer is ready to hand. As mentioned above, in the CAD model there are three different sorts of emotions, including, contempt, anger, and disgust. The violation of community norms causes contempt, the violation of autonomy norms arises anger, and the violation of divinity norms in secular societies spurs disgust. These norms not only can differentiate different kinds of negative emotions, and are also able to distinguish moral emotions and non-moral emotions. The contempt of the weakness of will in physical exercise is not a moral contempt because the contempt is not directed at the violation of community norms. The anger at another boxer is not a moral anger because the anger is not directed at the violation of autonomy norms. The disgust directed at excrement is not a moral disgust because the disgust is not direct at the violation of divinity norms. A negative emotion is a

moral negative emotion if and only if the negative emotion is directed at the violation of the norms in the CAD model.

Then the following question is: what makes the norms in the CAD model moral norms? A simple answer is that moral emotions make the norms moral norms. Another question is: what makes the emotions moral emotions? Another simple answer is that the moral norms make the emotions moral emotions. The mechanism is as follows:

> Imagine that certain behaviors cause emotions that are not yet specific to the moral domain. An act of cruelty might cause anger on the part of the victim, and sympathy among others. The perpetrator may be ostracized, criticized, and punished. This may cause the perpetrator to feel sad. If these responses are stable, then cruelty is governed by a kind of rule. The rule consists in the fact that cruelty is discouraged as a result of these emotional responses. The emotions guarantee a predictable pattern of behavior. Cruelty is less likely to occur, and when it does, certain emotions and corresponding behaviors will follow. After this pattern is established, the emotions that once had no moral significance take on new meaning. Sadness is not just a generic loss-response, but a feeling associated with violating a rule. Anger is not a generic response to a threat, but a feeling directed at rule violators. Guilt and righteous anger are born. At the very moment these emotions are born, the rule takes on new meaning. It is now a rule enforced by moral emotions. It is a moral rule. (Prinz, 2007, 118)

Some might worry that the creation of moral norms and moral emotions would be an example of vicious circularity (Prinz, 2007, 118). But I do not think that It is not a vicious circularity because this circularity is different the circularities we deem vicious. If we cannot explain A by B and then explain B by A because we believe that there is an ultimate explanation for the case. We think that there must be an explanation *before* other explanations. But we do not need to believe that there is a priority between moral norms and moral emotions. It is not either the case that moral norms precede moral emotions or that moral emotions precede moral norms

because they mutually depend on each other. We do not ask which organ in our body is more fundamental because the organs mutually rely on each other to keep the body alive. The heart has the function of pumping blood, the liver has the role of neutralizing toxins, the stomach has the function of digesting food, and so forth. When all the organs are placed together, they have a new task: keeping the body alive. The same goes for moral norms and moral emotions. Negative emotions have the functions of preventing a certain kind of behavior, and the norms the functions of guiding people to have a particular kind of negative emotions towards a certain kind of behavior. When the norms and the negative emotions are placed together, they have a new function: preventing people from violating the norms.

3 Concluding Remarks: Putting Things Together

Historically, sentimentalists understood moral judgments as the expressions of sentiments towards actions, but such understanding leads us to the problem of the recalcitrance of emotions. To avoid this difficulty, neo-sentimentalists understand moral judgment as the judgments of an appropriate emotion without actually having the emotion. The notion of being appropriate is analyzed in terms of reasons of the right kind. At this point, neo-sentimentalists unify reason and emotions in their framework. Meanwhile, in order to preserve the importance of emotions, it is argued that emotions are a more cognitively economical way to make moral evaluations. Emotions also make one sensitive to the values of the objects in environments and motivate one to act morally.

The bodily attitudinal theory of emotions illustrates how emotions bring one the sensitivity and moral motivation. Emotions are consciously felt physiological responses that consist of behavioral dispositions and other bodily phenomena. Since bodily attitudes are consciously felt, one is aware of the values of the objects in environments. Because emotions have behavioral dispositions, one is motivated to act morally while having the emotion, and such dispositions are as well implicit evaluations towards objects.

Neo-sentimentalism holds that to judge that an action is morally wrong is to judge that it is appropriate to resent killing, which means that we have a right reason to resent killing, without actually having the resentment. However, in order to be more morally competent, one also needs the help of emotions to sense moral values and to act morally.

Bibliography

- American Psychological Association (2012). *What You Need to Know About Willpower: The Psychological Science of Self-control*, URN = <http://www.apa.org/helpcenter/willpower.aspx>.
- Baker, D. (2014). The Abductive Case for Humeanism Over Quasi-Perceptual Theories of *Desire.* Journal of Ethics and Social Philosophy 8 (2): 1-29.
- Beauchamp, T. (1998). David Hume: An Enquiry Concerning the Principles of Morals: A Critical Edition. Oxford University Press.
- Bechara, A., Damasio H, and Damasio A. R. (2000). *Emotion, Decision Making and the Orbitofrontal Cortex*. Cerebral cortex 10.3 (2000): 295-307.
- Becker, K. (n.d). *Reliabilism*, Internet Encyclopedia of Philosophy, URN = < http://www.iep.utm.edu/reliabil/>
- Berthoz, S., Armony, J.L., Blair, R. J. R., Dolan, R. J. 2002. *An fMRI Study of Intentional and Unintentional (embarrassing) Violations of Social Norms*. Brain 125: 1696–708
- Birondo, N. (2014). *Aristole, Virtue, and the Wrong Kind of Reason*. In Dimensions of Moral Agency edited by David Boersema. Cambridge Scholars Publishing.
- Bittar, R. G., Otero, S., Carter, H., Aziz, T. Z. (2005). *Deep Brain Stimulation for Phantom Limb Pain. Journal of Clinical Neuroscience* 12 (4): 399–404.
- Blair, R. J. R. (1995). *A Cognitive Developmental Approach to Morality: Investigating the Psychopath*. Cognition 57: 1–29.
- Bonjour, L. (2002). *Intemalism and Extemalism*. In Paul K. Moser (ed.), The Oxford Handbook of Epistemology. Oxford University Press 234--263.
- Damasio, A. R. (2000). *The Feeling of What Happens: Body, Emotion and the Making of Consciousness*. London: Random House.
- Damasio, A. R. (2006). *Descartes Error: Emotion, Reason and the Human Brain.* London: Random House.
- D'Arms, J., Jacobson, D (2003). *VIII. The Significance of Recalcitrant Emotion*. Royal Institute of Philosophy Supplement 52: 127-145.
- D'Arms, J., Jacobson, D. (2000). *Sentiment and Value.* Ethics 110 (4):722-748.
- Davidson, D. (1987). *Knowing One's Own Mind.* Proceedings and Addresses of the American Philosophical Association 60 (3): 441-458.
- Deigh, J. (2010). *Concepts of Emotions in Modern Philosophy and Psychology*. In Peter Goldie (ed.), Oxford Handbook of Philosophy of Emotion. New York, Oxford University Press 17-40.
- Deonna, J. A. (2014). *In What Sense Emotions Are Evaluations?* URN = < http://www.affective-sciences.org/system/files/events/In%20what%20sense%20are%20emotions%20evaluations%20(II).pdf>.
- Deonna, J. A., Teroni, F. (2012). *The Emotions: A Philosophical Introduction*. Routledge.
- Finlay, S. (2007). *Four Faces of Moral Realism*. Philosophy Compass 2 (6): 820-849.
- Fisher, A. (2011*). Metaethics: An Introduction.* Routledge.
- Frijda, N. (1988). *The Laws of Emotions*. American Psychologist.

- Greene, D. (2007). *Why are VMPFC Patients More Utilitarian? A Dual-Process Theory of Moral Judgment Explains.* Trends in Cognitive Sciences 11 (8): 322-323.
- Giummarra, M. J., Gibson, S. J., Georgiou-Karistianis, N., Bradshaw, J. L. (2007). *Central Mechanisms in Phantom Limb Perception: The Past, Present and Future*. Brain Research Reviews 54.
- Goldman, A. (1979). *What is Justified Belief?* In George Pappas (ed.), Justification and Knowledge. Boston: D. Reidel 1-25 (1979).
- Greco, J. (2002). *Virtues in Epistemology*. In Paul Moser (ed.), Oxford Handbook of Epistemology. New York: Oxford University Press 287--315.
- Haidt, J., Bjorklund, F., Murphy, S. (2000). *Moral dumbfounding: When Intuition Finds No Reason.* Department of Psychology, Lund University.
- Hieronymi, P. (2005). *The Wrong Kind of Reason*. Journal of Philosophy 102 (9): 437 - 457.
- Holton, R. (2009). *Willing, Wanting, Waiting*. Oxford University Press.
- Huemer, M. (2005). *Ethical Intuitionism*. Palgrave Macmillan.
- Huemer, M. (n.d). *Phenomenal Conservativism*. Internet Encyclopedia of Philosophy
- James, W. (1884). *What is an Emotion?* Mind 9 (34):188-205.
- Johnson, G. (n.d). *Theories of Emotions.* Internet Encyclopedia of Philosophy, URN = < http://www.iep.utm.edu/emotion/>
- Klein, G. A. (1998). *Sources of Power*. Cambridge, MA: MIT Press.
- Koenigs, M., Kruepke, M., Zeier, J., Newman, P. (2011). *Utilitarian Moral Judgment in Psychopath.* Social Cognitive and Affective Neuroscience 7 (6): 708-714
- Lang, G. (2008). *The Right Kind of Solution to the Wrong Kind of Reason Problem*. Utilitas 20 (4):472-489.
- Levy, N. (2010). *Resisting "Weakness of the Will."* Philosophy and Phenomenological Research 82 (1): 134 - 155.
- McMillen, D. L., and Austin, J. B. (1971). *Effect of Positive Feedback on Compliance Following Transgression. Psychonomic Science*, 24: 59–61.
- Millikan, R. G. (2003). *Teleological Theories of Mental Content*. Encyclopedia of Cognitive Science, Vol. 11; Macmillan 2003, pp. 1138-1141.
- Millikan, R. G. (1996). *On Swampkinds*. Mind and Language 11 (1):103-17.
- Moll, J., Oliveirra-Souza, R. De, Eslinger P. J. (2003). *Morals and the Human Brain: A Working Model*. Neuroreport 14: 299–305.
- Neander, K. (2006). *Naturalistic Theories of Reference*. In: Michel Devitt and Richard Hanley (eds.), The Blackwell Guide to the Philosophy of Language. Malden: Blackwell Publishing, 374-392.
- Olson, J. (2009). *The Wrong Kind of Solution to the Wrong Kind of Reason Problem*. Utilitas 21 (2):225-232.
- Oppy, G. (n.d.) *Propositional attitudes*. Routledge Encyclopedia of Philosophy, URN = <https://www.rep.routledge.com/articles/propositional-attitudes>.
- Peters, U. (2014). *Teleosemantics, Swampman, and Strong Representationalism*. Grazer Philosophische Studien 90:273–288.
- Poston, T. (n.d.) *Internalism and Externalism in Epistemology*. Internet Encyclopeida of Philosophy, URN = <http://www.iep.utm.edu/int-ext/>.
- Prinz, J.J. (2006). *The Emotional Basis of Moral Judgments*. Philosophical Explorations 9 (1): 29-43.

- Prinz, J.J. (2015). *An Empirical Case for Motivational Internalism*. In Gunnar Björnsson, Caj Strandberg, Ragnar Francén Olinder, John Eriksson & Fredrik Björklund (eds.), Motivational Internalism. Oxford University Press 61-84.
- Prinz, J.J. (2007). *The Emotional Construction of Morals*. Oxford University Press.
- Prinz, J.J., Nichols, S. (2010). *Moral Emotions*. In John Michael Doris (ed.), The Moral Psychology Handbook. Oxford University Press 111.
- Rabinowicz W., Rønnow-Rasmussen, T. (2004). *The Strike of the Demon: On Fitting Pro-Attitudes and Value*. Ethics 114 (3): 391-423.
- Ramachandran, V. S., Hirstein, W. (1998). *The Perception of Phantom Limbs*. The D. O. Hebb lecture". Brain: a journal of neurology 121 (9): 1603–1630.
- Rosati, C. S. (2014). *Moral Motivation*. The Stanford Encyclopedia of Philosophy (Spring 2014 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2014/entries/moral-motivation/>.
- Rozin, P., Lowery, L., Imada, S., Haidt, J. (1999). *The CAD Triad Hypothesis: A Mapping between Three Moral Emotions (contempt, anger, disgust) and Three Moral Codes (community, autonomy, divinity)*. Journal of Personality and Social Psychology 76: 574–86.
- Samuelsson, L. *(2013). The Right Version of "the Right Kind of Solution to the Wrong Kind of Reason Problem." Utilitas* 25 (3):383-404.
- Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., Cohen J. D. (2003). *The Neural Basis of Economic Decision-making in the Ultimatum Game*. Science 300: 1755–58.
- Sauer, H. *(2011). The Appropriateness of Emotions: Moral Judgment, Moral Emotions, and the Conflation Problem*. Ethical Perspectives 18 (1):107-140.
- Scanlon, T. (1998). *What We Owe to Each Other*. Belknap Press of Harvard University Press.
- Schnall, S., Haidt, J., Clore, G. L., Jordan, A. H. (2008). *Disgust as Embodied Moral Judgment*. Personality & Social Psychology Bulletin, 34(8), 1096–1109.
- Schroeder, M. (2008). *Value and the Right Kind of Reason*.
- Schroeder, M. (2008). *How Does the Good Appear to Us?* Social Theory and Practice 34 (1): 119-130.
- Shafer-Landau, R. (2003). *Moral Realism: A Defence*. Oxford University Press.
- Slote, M. (2003). *Sentimentalist Virtue and Moral Judgement: Outline of a Project*. Metaphilosophy 34 (1-2):131-143.
- Slote, M. (2010). *Moral Sentimentalism*. Oxford University Press.
- Smith, J. (2014). *Are Emotions Embodied Evaluative Attitudes?* Disputatio 6 (38):93-106.
- Sripada, C. S. (2010). *Philosophical Questions About the Nature of Willpower*. Philosophy Compass 5 (9):793–805.
- Watson, L. (2011). *Comments on Michael Slote's* Moral Sentimentalism. Southern Journal of Philosophy 49 (s1):142-147.
- Wheatley, T., Haidt, J. (2005*). Hypnotically Induced Disgust Makes Moral Judgments More Severe.* Psychological Science 16: 780–84.