Master Degree Project

1977

UNIVERSITY OF SKÖVDE

MORAL INTUITION VERSUS MORAL REASONING IN THE BRAIN

Master Degree Project in Cognitive Neuroscience
One year Level 30 ECTS
Spring term Year 2014

Andreas Ljungström

Supervisor: Stefan Berglund
Examiner: Judith Annett

MORAL INTUITION VERSUS MORAL REASONING IN THE BRAIN

Submitted by Andreas Ljungström to the University of Skövde as a final year project towards the degree of B.Sc. in the School of Bioscience. The project has been supervised by Stefan Berglund.

Date 11/6-14

I hereby certify that all material in this final year project which is not my own work has been identified and that no work is included for which a degree has already been conferred on me.

Signature: ____

Moral Intuition Versus Moral Reasoning In the Brain.

Andreas Ljungström

School of Bioscience

University of Skövde, Sweden

Abstract

Humans express complex moral behaviour, from altruism to antisocial acts. The investigation of the neural and cognitive mechanisms underlying our moral minds is of profound importance for understanding these behaviours. By reviewing recent findings in cognitive and moral neuroscience, along with other relevant areas of research, the current study aims to: (1) Investigate the neural correlates of moral intuition and moral reasoning, and see how these two systems relate to moral judgement and moral behaviour. (2) Examine how the moral intuitive system and the moral reasoning system relate to one another. Neuroscientific evidence suggests that these two systems are supported by different areas in the brain. While their relationship is argued to be both sequential, integrative and competitive, evidence indicates that the moral reasoning system primarily functions as a post hoc rationalization of our intuitive-driven judgements and behaviours. While our moral intuitive system motivates kin altruism, both moral intuition and moral reasoning serve to uphold reciprocal altruism.

Keywords: Morality, Moral Intuition, Moral Reasoning, Moral Judgement, Moral Behaviour.

Table of content

Introduction

Most human beings have probably experienced that the mind sometimes pulls us in different directions. The fact that the mind wants and desires many things, and that these needs sometimes conflict is a common view in psychology (Haidt, 2012). Many ancient thinkers have given us metaphors to understand people's inner battles. The roman poet Ovid had one of his characters in his work Metamorphoses plain:

> "But against my will some force bewitches me; One way desire, another reason calls; The better course I see and do approve - the worse I follow". (Ovid, 1986, p. 144)

Surely many people can identify with the character's ambivalence. Many times we know what we should do, or what is the "right thing to do", but somehow we end up doing the contrary. Many writers, poets and philosophers have written about the battle of reason versus emotions and they seem to emphasize two different things: what the relationship *is*, and what *should be*.

In Medieval Europe, Christian philosophers denigrated the role of the emotions since they were linked to desire and hence to sin (Haidt, 2001). The emphasis was on what we *should* do, that is, resist the many "sinful" temptations we have within us. This degradation of emotions continued with the 17th century's famous rationalists (e.g., Leibniz, Descartes) who worshiped reason and tried to build a philosophy based on the deductive method (Haidt, 2001). However, in the 18th century English and Scottish philosophers started to discuss alternatives to rationalism. They argued for a built in moral sense in humans that creates pleasurable feelings towards good and benevolent acts and feelings of disapproval toward evil acts (Haidt, 2001). One of the front figures of these theories was David Hume, who wrote in 1739:

> "Reason is, and ought only to be, the slave of the passions, and can never pretend to any other office than to serve and obey them" (Hume, 1969, p. 462).

For Hume, the role of reason in action was limited to the discernment of the means to our ends. He claimed that our ends were picked out by our passions, and reason was only a servant finding whatever means to achieve what the passion wanted (Hume, 1969).

Reason, as he saw it, could never be a motive to any action of the will, and it could never oppose passion in the direction of the will. In the moral domain, good and evil are distinguished by our sentiments, not by reason (Hume, 1969). For example, reason could tell us that a certain act would kill many innocent people, but unless we have some sentiment that value human life, reason cannot advise against taking the action. Instead of seeing morality as something transcendent that emerged from the nature of rationality, like Plato and other enlightenment thinkers did, Hume saw sentiments as the foundation of morals and the central force behind our actions.

Hume's sentimentalist approach to ethics was not well received by other philosophers. One who tried to refute Hume was the 18th century philosopher Immanuel Kant. While Hume argued that sentiments were the foundation of morality, Kant argued that ethics had to be based on reason alone. For him, pure rationality characterized moral obligation and in order to act truly moral, we had to detach ourselves from emotions and act simply on duty (Kant, 2002). Emotions and passions were often obstacles to rational self-control and they could without the good will control become extremely evil (Kant, 2002). He further argued that we should only act on principles that could be made into a universal law (Kant, 2002). To act on duty and good will was for him an end in itself, no matter the consequences.

In the late 19th century however, psychologists seemed to tone down the role of reason as a guiding force of our behaviour. Freud and other psychoanalysts introduced the world to the ideas of an unconscious mind motivating our behaviour by a combination of repressed emotions and innate drives, and with a conscious mind prone to rationalizations and self deception (Evans, 2008). Until the cognitive revolution in the 1960's, psychologists did not give much praise to reason. Instead their views on morality were more compatible with Hume's sentimentalist theory (Haidt, 2001). However, in the 1970's, the emphasis on emotions and unconscious forces guiding our behaviour was challenged by Lawrence Kohlberg, who's work played an important part in the cognitive revolution (Haidt, 2001). Building on previous work by among others Plato, Hegel and Piaget, Kohlberg presented a cognitive development theory, arguing that children progress in six stages when reasoning about the social world (Kohlberg & Gilligan, 1971). In early development, children are mostly concerned with their own needs and they follow rules for its own sake. However, as their cognitive abilities develop, they start understanding that the ideas behind laws and personal rights serve the end of social welfare. In the last stage, which is the end goal, one understand that right is defined by the decision of conscience in accord with self-chosen ethical principles that appeal to logical comprehensiveness, universality and consistency (Kohlberg & Gilligan, 1971). For Kohlberg, morality was about justice, reciprocity and equality of human rights, and the moral force in personality was cognitive. Once again, rationality, logic and abstract thinking ruled. However, a few decenniums later, social psychologist Jonathan Haidt (2001) challenged Kohlberg and his cognitive development theory. He wondered whether reasoning was really the cause of a moral judgment, rather than a consequence. If reason was the force that guided people's moral judgement, and if morality was about harm and fairness, then people should find nothing immoral about stories that were offensive but contained neither harm or injustice. He came up with stories that were offensive yet harmless, such as eating one's dead pet dog, cleaning one's toilet with the national flag or eating a chicken carcass one just had used for masturbation. Even though the stories were made up so that no plausible harm could be found, most people still said that the actions were wrong, and sometimes universally wrong. Further more, their affective reactions to the stories were better predictors of their moral judgements than their claims about harmful consequences (Haidt, 2001). With these findings he presented the Social Intuitionist Model (Haidt, 2001) in which the basic claim was that moral judgement is caused by quick moral intuitions followed by (when needed) slow ex post facto moral reasoning. Haidt (2001) claimed that the relationship between the two was that reason is more like a lawyer defending a client (the intuitions) than a scientist seeking the truth.

Several things are important to point out here. First of all, it seems to be more to morality than harm and fairness, something that will be discussed later on. Second, the battle between reason and intuition appears to be constituted by two distinct processes in the brain: one fast, affective and automatic (intuitions) and one slow, conscious and deliberate (reasoning). Many other scientists have recognized the dual process system in cognition. While Haidt refers to the systems as intuition and reasoning, they are normally also referred to as implicit and explicit cognition (Greenwald & Banaji, 1995) or to System 1 and System 2 (Kahneman, 2011). What these dual process theories have in common is the idea that there are two distinct modes of processing in the brain that contrast each other. The first one (System 1) is rapid, unconscious and automatic, and the other one (System 2) is conscious, slow and deliberative (Evans, 2008). Even though Haidt (2001) made a strong claim about their sequential relationship in moral judgement, it is still under debate whether it is sequential or parallel (Evans, 2008).

The developing field of moral cognitive neuroscience has made important contribution towards a better understanding of the two system's roles in moral judgements and moral behaviour. Neurologist Antonio Damasio has provided interesting evidence for the role of the intuitive system in moral judgment and behaviour through his studies on patients with damages to the prefrontal cortex (PFC). In his book Descartes' error (1994) he presented his hypothesis of somatic markers, claiming that our experiences often elicit an emotional experience that most importantly expresses itself through changes in the representation of the body state. He designated these emotional changes under the umbrella term "somatic states". This emotional component helps us make decisions fast and automatic. The ventromedial prefrontal cortex (VMPFC) is a part of the frontal lobes where emotional responses are integrated with a person's other knowledge and planning functions. This integration is crucial in order to decide quickly on a response to an event (Damasio, 1994). In studies of patients with damage to the VMPFC, Damasio (1994) saw a consistent pattern. The patients performed well on standard cognitive tests, such as IQ tests, but made disastrous decisions in real life. They seemed to have a reduced ability in making judgements and decisions on the personal and social level due to their emotional deficits, despite their intact reason ability (Damasio, 1994).

This new emphasis on emotional processes and their effect on judgement and decision making inspired philosopher and neuroscientist Joshua Greene. He started doing research with the goal of understanding how moral judgements were shaped by automatic processes (emotion) and controlled cognitive processes, such as reasoning and self-control.

Primarily using functional magnetic imagine (fMRI) and behavioural experiments, he presented people with different moral dilemmas called "personal" and "impersonal" (Greene, Sommerville, Nystrom, Darley & Cohen, 2001). An impersonal dilemma could be to approve or disapprove of diverting a runaway trolley that mortally threatens five people onto a side-track, where it would kill only one person. A personal dilemma could be to push someone in front of a runaway trolley, which would kill the person being pushed, but save five others (Greene, 2007). The puzzle has been that people generally approve of diverting a runaway trolley through a switch (impersonal dilemma) and thereby sacrifice one to save five. However, people generally disapprove of pushing a man in front of a trolley (personal dilemma) in order to save five (Greene, 2007). With his research he explained this phenomenon by referring to a dual process system that is involved in moral reasoning: one automatic emotional system and one cognitive reasoning system. The personal dilemma seems to trigger the automatic system, which creates a negative emotional response that encourages moral disapproval. However, in situations where there is no pre-potent emotional response, such as in some impersonal dilemmas, the cognitive reasoning system seems to prevail and produce a more utilitarian moral judgement (Greene, 2007). Supporting this theory are findings from studies on people with emotional deficits, such as patients with damages to the VMPFC. In the personal dilemmas, these patients were more likely to give utilitarian answers. For example, they were more likely of pushing a man in front of the trolley in order to save five people (Greene, 2007) or taking an action that would result in killing their own family members in order to save a larger number of strangers (Greene, 2013). Greene's work supports Haidt's claim (2001) that there is a dual process involved in moral cognition, one affective and automatic and one deliberate and "rational". However, they seem to disagree on certain aspects. While Haidt (2001) claims that reasoned judgement within an individual is rare and only occurs when intuitions are weak and processing capacities are high, Greene's dual process model allows that moral reasoning, especially utilitarian/consequentialist reasoning, may be a common feature of human moral sense (Paxton & Greene, 2010). If our intuitive system is in charge, as Haidt (2001) believes, then the only way to change other people's judgement and behaviour is to modify their affective responses. However, Greene seems to give reason a slightly more influential role, claiming that people can influence other's judgements and behaviour by transmitting moral principles that can be used to override our affective responses (Paxton & Greene, 2010).

Purpose/Aim of the Essay

This new emphasis on intuitions and its affective components has been the target for much of the emerging field of social neuroscience. One important breakthrough has been the growing understanding that the intuitive system appears to play a central role in human social behaviour (Moll, de Oliveria-Souza & Zahn, 2008; Damasio, 1994). Neuroscience and its new technology to measure brain activity has provided new insights to the neural basis of human moral cognition (Moll, Zahn, de Oliveira-Souza, Krueger & Grafman, 2005). Along with the growing understanding of the importance of the intuitive system in moral cognition, Moll et al. (2005b) argues that one of the central aims of cognitive neuroscience is to clarify the neural basis of moral emotion and how they relate to moral cognition. Further, the relationship between the intuitive system and the reasoning system in moral cognition is also poorly understood (Moll & Schulkin, 2009; Evans, 2008; Mallon & Nichole, 2011; Paxton & Greene, 2010). The aim of this essay is to review recent research in the field of cognitive neuroscience, along with other relevant areas of research, in order to:

(1) Investigate the neural correlates of moral intuition and moral reasoning, and see how these two systems relate to moral judgement and moral behaviour.

(2) Examine how the moral intuitive system and the moral reasoning system relate to one another.

Method and Limitations

One of the major challenges for moral cognitive neuroscience is that it requires extensive cross-field integration of neuroscience, philosophy, psychology, evolutionary biology and anthropology among other areas (Moll et al., 2005b). Neuroscience plays a crucial part in the search for, and understanding of the neural basis of moral cognition. Psychology and social psychology are important in the understanding of moral cognition and context dependent judgement and behaviour. Since the general consensus today is that the building blocks of morality have strong evolutionary bases, evolutionary biology also play an important role in the understanding of certain primitive motivational-emotional mechanisms identified in both humans and other species (Moll & Schulkin, 2009). In order to answer the questions for this essay, all of these areas will be considered. However, in the domain of cognitive neuroscience, I will mainly focus on studies conducted between the years 2000-2014 by three influential scientists: Joshua Greene, Jonathan Haidt and Jorge Moll (with colleges). These three have done extensive work in the field of cognitive moral neuroscience over the past 15 years. Further, they all seem to have slightly different theories concerning the relationship between the moral intuitive system and the moral reasoning system. I will use their different approaches and compare them in order to find strengths and weaknesses in each theory.

In the first part morality and its foundation will be discussed from an evolutionary and neuroscientific perspective. Thereafter morality, moral judgement and moral behaviour will be defined. I will then continue to examine the intuitive system in the brain, its neural correlates and its effects on moral judgement and behaviour. Thereafter, the reasoning system will be examined, along with its neural correlates and its effects on moral judgement and behaviour. Finally I will discuss my findings in relation to the purpose of the essay.

Morality and Cooperation

Many scientists (e.g. Greene, 2013; Churchland, 2011; Haidt, 2008) argue that morality emerges from a coevolution of psychological mechanisms and cultural innovations in order for selfish individuals to profit from cooperation. From an evolutionary perspective, cooperation could never have evolved if it didn't confer a competitive advantage on the co-operators. However, cooperation is dependent on unselfishness and that individuals are willing to pay personal costs in order to benefit the group (Greene, 2013). Haidt (2008) sees moral systems as " interlocking sets of values, practices, institutions, and evolved psychological mechanisms that work together to suppress or regulate selfishness and make social life possible" (p. 70). Greene (2013) takes a similar approach, defining morality as "a set of psychological mechanisms that allows otherwise selfish individuals to reap the benefits of cooperation" (p. 23). The emphasis is on unselfishness, which apparently is a crucial ingredient for cooperation to thrive. If unselfish (or altruistic) behaviour is defined as behaviour which benefit others at some cost to oneself, altruism in non-humans is well documented (Singer, 2011). Some mammals can risk their lives protecting their cubs or other members of their specie. Many animals also share food within, and sometimes outside, the group and occasionally help wounded members of the flock (Singer, 2011). From an evolutionary perspective, socio-biologists suggest that two types of altruism can be explained in terms of natural selection: kin altruism and reciprocal altruism. Kin altruism builds on the idea that evolution can be regarded as a competition for survival among genes, and this can only be done by successful reproduction. Thus, strictly selfish behaviour would not be favoured by evolution (Singer, 2011). Caring for one's own well-being is a ground-floor function of nervous systems and brains are organised to seek well-being and relief from ill-being (Churchland, 2011). However, the crucial step from self-caring to other-caring, typical of mammals, depend on the neural-and-body mechanism that "maternalize" the female mammal brain which in turn depend on the hormones oxytocin and vasopressin (Churchland, 2011). Thus many mammals care about their pups well-being in the same way that they care about their own well-being, and this function is believed to have evolved since it gave children of caring parents a better chance to survive. The ability to extend the circle of care is crucial in altruistic behaviour and hence a cornerstone in cooperation. However, not all altruistic acts help relatives but also other members of the group or specie. For example, monkeys grooming each other are not always related. Here reciprocal altruism offers an explanation: you scratch my back and I'll scratch yours (Singer, 2011). In order for reciprocal altruism to work, the group cannot have members who gladly take but never reciprocate. However, abilities to recognize and remember who is a good co-operator and who is not, and to punish people who cheat, make it possible for reciprocal altruism to work. These abilities are to different degrees seen in social animals and they are especially high in more intelligent social animals like wolves, wild dogs, dolphins, baboons, chimpanzees and human beings (Singer, 2011).

Another important ability important for cooperation and sociality is the ability to make good predictions in the social sphere. Primates have a larger prefrontal cortex than most animals, which creates a more complex and flexible connection between stimuli and response. Humans are in some sense smarter than other mammals since we have a greater cognitive flexibility, a greater capacity for abstraction and long-term planning and a strong ability to imitate (Robbins & Arnsten, 2009). Making good prediction helps us not only to see what other people will do or to what extent we can trust them as good cooperation partners, but they also enable us to adjust our own behaviour in order to reduce or avoid conflict (Churchland, 2011).

The last important ability many social animals share is that we can learn social practice systems. Our emotional punish and reward system, together with our ability to imitate, is crucial in the learning of social practices, learned altruism and other forms of conditioning (Churchland, 2011; de Waal, 2008).

In sum, cooperation in humans and in other animals has evolved because of its evolutionary advantage, and morality is nature's way of dealing with the tension between one owns needs and other's needs (Greene, 2013). Several capacities enable us to have effective cooperation: (1) urges to care about the welfare of self, mate, offspring and affiliates, which motivate altruistic behaviour; (2) the capacity to evaluate and predict what oneself and others will feel or do in particular circumstances; (3) a neural reward-punishment system that is linked to internalizing social practices and applying them suitably (Graybiel, 2008). Although these capacities enable cooperation, there is more than one way to handle the problem of free-riders and encourage selfless acts. A common approach in the western tradition is an individualistic approach to morality where the conception of persons as reasoning beings with equal worth who must always be treated as ends in themselves and never solely as means to other goals (Haidt, 2008). According to Haidt, (2008) this tradition puts much emphasis on individual rights and our justice system reflects a liberal narrative more than a community narrative. Selfishness is suppressed by encouraging individuals to care for the needy and vulnerable and to respect other's fight for justice. Authority and tradition have no value in and of themselves, instead they should be treated with scepticism and be ready to change in order to suit the societies changing needs (Haidt, 2008). However, not all cultures embrace an individualistic approach to morality. Although issues related to harm and fairness appear to be found in all cultures (Hauser, 2006), issues related to in-group loyalty, authority, respect and spiritual purity are important in many non- western cultures (Haidt, Koller & Dias, 1993).

In these cultures the group is treated as the highest source of value, not the individual. Haidt (2008) calls this a *binding* approach to morality because these cultures bind people into larger collectives like families, teams or congregations and they expect the individual to limit their desire and play their roles within the group. Hence when studying morality, it is important not to narrow it down into concepts about harm and fairness, but also embrace values from non-western cultures like authority, respect, purity, sanctity and loyalty. There are different ways of supressing selfishness and when one realizes this, it enables to study morality in its many forms and expressions.

Defining Morality, Moral Judgement and Moral Behaviour

Definitions are important tools but they can also blind. As previously argued, morality can be seen as an expression of social animal's attempt to regulate selfishness in order to profit from cooperation. Since different groups, cultures and societies use diverse strategies to do this, many acts and judgements will inevitably fall under the definition of morality. One way to set up a limit would be to claim that morality first and foremost is about harm and fairness, hence behaviour and judgement that relates to these two would then be considered moral or immoral. However, this approach tends to leave out the possibility that there are many other important strategies of supressing selfishness used in other societies and cultures. Haidt (2008) argues that moral systems are " interlocking sets of values, practices, institutions and evolved psychological mechanisms that work together to suppress or regulate selfishness and make social life possible" (p. 70). This definition is open to the possibility of many strategies serving the same end, and it puts emphasis on how psychological mechanisms and existing values and practices work together on creating, maintaining or changing these strategies. The aim of this essay is obviously not to have a normative discussion of what is right and wrong. Rather it is about understanding the psychological mechanisms that underlie people's statements concerning their subjective experience of right and wrong. Haidt's definition therefore serves this purpose well. Following this definition of morality, moral judgement will be defined as an evaluation (good vs. bad) of a person's actions or character that are made in relation to a culture's set of customs, values and virtues (Haidt, 2001). Moral behaviour will similarly be defined as an action or behaviour that corresponds to, or violates customs, values or virtues held by a culture or a subculture.

Moral Intuition

> Julie and Mark are brother and sister. They are travelling together in France on summer vacation from college. One night they are staying alone in a cabin near the beach. They decide that it would be interesting and fun if they tried making love. At the very least it would be a new experience for each of them. Julie was already taking birth - control pills, but Mark uses a condom

> too, just to be safe. They both enjoy making love, but they decide not to do it again. They keep that night as a special secret, which makes them feel even closer to each other. What do you think about that? Was it OK for them to make love?

This example is presented by Jonathan Haidt (2001) and when asking people what they think of the story, most people say that what Julie and Mark did was wrong. After the participant's initial judgement, they started to search for reasons (Haidt, 2001). The story was carefully made up so that no plausibly harm could be found. Even though people could not provide good reason for why they thought the act was wrong, they still would not change their initial judgement (Haidt, 2001). What model of moral judgement can explain why people sometimes have a strong sense of what is right and wrong without knowing why? The answer to this question might come from theories concerning moral intuitions.

Moral intuitions are by contemporary philosophers seen as a moral assessment or judgement that occur quickly and automatically and carry with them a strong feeling of authority or appropriateness without having to go through any conscious reasoning process that leads to this assessment (Woodward & Allman, 2007). A broadly similar understanding of moral intuitions is held among many psychologists and neurobiologists (Woodward & Allman, 2007). Haidt (2001) defines moral intuition as "the sudden appearance in consciousness of a moral judgment, including affective valence (good–bad, like–dislike) without any conscious awareness of having gone through steps of search, weighing evidence or inferring a conclusion". What differs Haidt's view of intuition from some philosopher's views is the affective component in intuitions. However, although some philosophers dismiss or downplay the role of emotion and affect in moral intuition, there is considerable empirical evidence that the neural areas involved in moral intuition are also centrally involved in emotional processing. Manipulation of emotional processing has moreover shown to affect the content of one's moral intuitions (Woodward & Allman, 2007). Haidt's (2001) definition will therefore be used and central to the definition will be the affective component in intuitions. Moral reasoning stands in contrast to intuitive moral processes since it is a conscious, more deliberative process that consists of transforming given information about people in order to reach a moral judgement (Haidt, 2001).

Moral intuitions can be seen as a form of social cognition, which has to do with information processing in the social world (Woodward & Allman, 2007). When navigating in the social world, we need to be able to predict and understand the behaviour and mental states of others, recognize behaviour and intentions that are potentially beneficial or harmful to oneself or those one cares about and respond appropriately to others behaviours and intentions. We also need the ability to understand and anticipate how others are likely to respond to one's own choices and so on (Woodward & Allman, 2007). Social cognition and social behaviour has commonly been treated as operating under conscious and thoughtful control (Greenwald & Banadji, 1995). However, evidence supports the view that social cognition often operates in an implicit or unconscious fashion (Greenwald & Banadji, 1995). Many social situations are extremely complex, and if we only were to trust our rational system to guide our judgements and decisions, calculations of wins and losses in relation to available knowledge would take an extremely long time (Damasio, 1994). Further, our attention and our working memory have limited capacities and cannot store and remember all the information in complex social interactions that would be relevant in order to make wise decisions. Instead we are born with a predisposed affective machinery that throughout life evolve and connect affective responses to stimuli and experiences in order to help us make wise and quick decisions, especially in the social sphere (Damasio, 1994). This connection is mainly mediated through the VMPFC, and damages to this part of the brain have shown to be potentially catastrophic to people's social lives (Damasio, 1994). According to Moll et al. (2008), Haidt (2001) and Woodward and Allman (2007), moral intuition (with its emotional component) appears to play a central role in moral cognition. Moll et al. (2008) distinguishes moral cognition from other socially relevant abilities by its tendency to altruistically motivate behaviour. And what appears to play a central role in motivating altruistic behaviour is a set of emotions called the moral emotions. These emotions are, in my interpretation of Haidt, central to moral intuition because they provide the affective valence (good-bad, like-dislike) that many times guide our moral intuitions.

The moral emotions

Moral emotions are, as previously argued, central to moral intuition and to moral cognition. They appear fast and automatic with a conscious awareness of the outputs, but not of the process (Haidt, 2012). Moral emotions are linked to social contexts and readily evoked by the perception of moral violations (Moll et al., 2002). A similar view is held by Prinz (2007), who claims that moral emotions promote or detect conduct that violates or conforms to a moral rule. Haidt (2003) defines moral emotions as "those emotions that are linked to the interest or welfare of either society as a whole or at least some other person other than the judge or agent". This definition fits well with Moll's (et al, 2008) argument that moral emotions are central to moral cognition in that they motivate altruistic behaviour.

According to Haidt (2003), a prototypical moral emotion has two typical features: disinterested elicitors and pro-social action tendencies. Many emotions arise when good or bad things happen to the self but in some cases emotion arises when we witness or hear about events that has no direct effect on the self. For example, seeing a child suffer can trigger sympathy while hearing about an unjust event can make us feel anger. The more an emotion is triggered by such disinterested events, the more it can be considered a moral emotion (Haidt, 2003). A prototypical moral emotion also has pro-social action tendencies. An emotion generally motivates some form of action, and these action tendencies can be ranked by the degree they benefit others or hold up social order (Haidt, 2003). The more pro-social action tendencies an emotion has, the more it is considered a prototypical moral emotion.

Moral emotions are typically divided into four different categories; the other-condemning emotions (disgust, anger, contempt), the self-conscious emotions (shame, guilt, embarrassment), the other-suffering emotions (empathy, compassion) and the other-praising emotions (gratitude, elevation) (Haidt, 2003). They will all be discussed in terms of elicitors and how they affect moral judgment and moral behaviour.

The other-condemning emotions.

Human beings live in a rich moral world of reputations and third party concerns (Haidt, 2003). Through language and highly developed social-cognitive abilities human beings are allowed to keep track of the reputation of hundreds of individuals (Dunbar, 1996). Singer (2011) argues that reputation is an important factor in reciprocal altruism. If we help someone that does not return the favour, we can cease to help that person in the future. However, if we tell people around us what a poor co-operator he/she was, then other people also might be less likely to help that person in the future. We care about what other people do so much that we spend hundreds of hours gossiping in order to catch cheaters, liars, hypocrites and other who try to fake the appearance of reliable cooperation partners (Haidt, 2003). Moral violators often become the subjects of our negative "other-condemning" emotions, which are contempt, anger and disgust (Haidt, 2003).

Disgust. Disgust has been recognized as a basic emotion that has a characteristic facial expression, an appropriate action (distancing of the self from an offensive object), a distinctive psychological manifestation (nausea) and a characteristic feeling state (revulsion) (Rozin & Fallon, 1987). It has been connected to activation in occipito temporal regions, prefrontal brain structures and the amygdala (Stark et al., 2007). Even though the amygdala and the insula most strongly appears to correlate with disgust rates, the exact role of the insula within this network remains unclear (Stark et al., 2007).

The word disgust means "bad taste" and it is believed to have evolved to help omnivorous species decide what to eat in a world full of microbes and parasites that spread by physical contact (Rozin & Fallon, 1987). Although disgust evolved as a food related emotion that helped organisms avoid or expel a substance, it was also well suited as an emotion of social rejection (Schnall, Haidt, Clore & Jordan 2008). In many cultures, the word and facial expression use to reject physically disgusting objects are also used to reject certain kinds of socially inappropriate people and behaviour (Haidt, Rozin, McCauley & Imada, 1997).

Moral disgust is often divided into two types: one that is elicited by moral violations to the body (e.g., cannibalism, paedophilia and incest) and one elicited by moral violations that does not involve the body (e.g., deception, betrayal and offensive attitudes) (Russell & Rogers, 2013; Haidt et al., 1997). Researchers still debate whether different elicitors evoke the same form of disgust or if moral disgust is more connected to anger and indignation (Russell & Rogers, 2013: Moll et al, 2005a). Moll et al. (2005a) tested whether moral disgust had different activation patterns than non-moral disgust, but found remarkable overlapping neural substrates. However, the orbital frontal cortex (OFC) and the insula were more activated in moral disgust than in non-moral disgust. The results had the authors conclude that the different part of the OFC might be responsible for producing different subtypes of disgust that are evoked in social contexts (Moll et al, 2005a). However, the difference in neural substrates between moral disgust elicited by violations to the body and moral disgust elicited by violations that does not involve the body remains unclear.

Haidt et al. (1997) argue that bodily disgust can be seen as a guardian of the temple of the body, or as a defence of the distinction between animals and humans. We fear our animal nature because it reminds us of our mortality. This type of disgust is weakly associated with situation appraisals, such as whether the behaviour is harmful or justified. Some objects or acts just seem disgusting to people, even though they cannot always give good reasons why. Bodily disgust can therefore be seen as a more unreasoned emotion (Russell & Rogers, 2013).

The other form of disgust is thought to be elicited by violations against norms about fairness, harm or rights. This reported form of moral disgust tends to co-occur with anger, and is not as unreasoning as bodily moral disgust is (Russell & Rogers, 2013). Even though some argue that this form of disgust might be mistaken for anger, it is still debateable whether they co-occur or if they are confused with each other (Russell & Rogers, 2013).

More than any other emotion, disgust might be our strongest "gut feeling" because of its link to nausea. This bodily reaction can be so well learned and associated to certain stimuli or events that whenever people merely think about a similar situation, the same feeling can be experienced (Schnall et al., 2008). Recent experimental research has tried to reveal the relationship between disgust and morality. Wheatley and Haidt (2005) found that hypnotically induced disgust results in more severe *moral judgements* of both morally and non-morally relevant behaviour. In another series of studies by Shnall et al. (2008), they induced disgust in the participants via an unpleasant odour, a dirty environment, a movie clip and a writing task. The studies found that people that were sensitive to internal sensations (high private body consciousness) made more severe *moral judgement* in moral relevant behaviours when disgust was induced (Shnall et al., 2008). Although induced disgust appears to have an effect on moral judgement and that some people seems more sensitive to disgust, it is not clear if the casual link between experienced disgust and moral decision making applies for all individuals (David & Olatunji, 2011).

In *moral behaviour*, all forms of disgust motivate people to avoid, expel or break off contact with the offending entity (Haidt, 2003). Disgust also motivates people to wash, purify or otherwise remove residues of any physical contact with the offending entity (Rozin & Fallon, 1987). For example, books presenting disgusting ideas can be labelled as "filth", banned from libraries and in extreme cases burned (Haidt, 2003). In society we ostracize people that trigger moral disgust by setting up punish and reward systems that tries to deter culturally inappropriate behaviour, especially those involving the body (Haidt, 2003).

Disgust is also believed to interfere with the ability to mentalize with the disgusting object. Mentalizing refer to the process by which we make inference about mental states, or "reading other's minds". A brain area crucial to this ability is the medial PFC (Frith & Frith, 2006). In an fMRI study by Harris & Fiske (2007) participants viewed images of stigmatized groups typically associated with disgust, such as homeless people and drug addicts. Viewing images of these groups reduced activation the medial PFC, which suggests that people might be disinclined to make inferences about the mental states of people belonging to stigmatized groups (Harris & Fiske, 2007). Feelings of disgust towards others in combination with a low motivation to emphasise with them might lead to dehumanisation processes where we experience and treat people as less human (Sherman & Haidt, 2011). In a special form of dehumanization, called animalistic dehumanization, the target is perceived as crude, savage and similar to non-human animals. This form of dehumanization is strongly linked to disgust and directly shrinks the moral circle, excluding outgroup members from our moral concern (Sherman & Haidt, 2011). Some researchers also speculate whether impairment in recognizing disgust might be correlated with sexual deviant behaviour, such as paedophilia (Moll et al, 2005a).

Anger. Anger is a negative valenced, other-focused emotion that is not typically considered in the moral sphere (Tangney, Stuewig & Mashek, 2007). People may experience anger in a broad range of situations – e.g., when frustrated, insulted, inconvenienced or injured in any other way (Tangney et al., 2007). Batson et al. (2011) identifies and distinguishes three different types of anger: personal anger, empathic anger and moral outrage. Personal anger is often elicited when one's own interests has been thwarted or blocked. The appraisal evoking empathic anger is that the interests of a person for whom one cares for has been thwarted. Moral outrage arises in response to a special class of anger eliciting events, those in which the perpetrator's behaviour is seen as a violation of moral standards (Tangney et al., 2007). Batson, Shaow & Givens (2009) however, encourage caution when it comes to categorize anger as a moral emotion. In their study they compared empathic anger and moral outrage, and found little evidence that moral outrage in response to violation of a moral standard actually existed. Their argument was that what makes people angry is not the violation itself, but instead the harm it does to others. For example, they presented the participants with two different stories about torture. In the first story, a marine from the USA, with whom the participant's shared national identity, was tortured. In the second story a Sri Lankan soldier, with whom the participants didn't share a national identity, was tortured. Even though the violation in the stories where the same, the first story produced much more anger in the participant's than in the second story (Batson et al., 2009). So should we dismiss moral outrage as a moral emotion? Seemingly moral outrage is heavily dependent on our relation to the victim(s), therefor maybe empathic anger would be more appropriate in the identification of moral emotions, and not moral outrage.

Unfairness and immorality have been shown to be two strong elicitors of anger (Scherer, 1997). However, one can question if anger is elicited by the violation itself or by the blockage of goals or interests the violation produces for ourselves or significant others. Empathic anger's effect on *moral judgement* seems, as previously argued, dependent on our relationship to the victim. Consider the example by Batson's et al. (2009) study with the two soldiers with different nationalities being tortured. In both cases, the participants judged the action to be wrong while the story about the soldier that had the same nationality as the participants elicited a higher degree of anger. While we may not be dependent so much on empathic anger in order to judge an action to be unjust or unfair, our choice to act morally, or pro-social, will be determined by to what extent our action helps us to reach egoistic or altruistic goals.

The neural basis of anger is not yet fully understood. A positron emission tomography (PET) study by Dougherty et al. (1999) examined the neuroanatomy of anger in 8 healthy men. They used narrative scripts to induce anger and measured normalized regional cerebra blod flow (rCBF). Statistical parametric maps where constructed to reflect the anger versus neutral state contrast. The result showed that anger was associated with activation in the left orbitofrontal cortex, right anterior cingulated cortex affective division and bilateral anterior temporal poles (Dougherty et al., 1999). However, the results should be treated with caution. State induction experiments are often by their nature biased by collateral affective and cognitive processes. Secondly, there were only 8 participants in the study, and all of them were men. Questions that need further explanation is whether the neuroatonomy of anger is the same in women, if anger and empathic anger share the same neural correlates and if a study of 8 men is really enough to draw general conclusions over a larger population.

Contempt. Contempt has been recognized as a basic emotion (Ekman, 1992), but it differs from disgust and anger in that it does not have a clear animal origin (Rozin, Lowery, Imada & Haidt, 1999). Like the moral forms of anger and disgust, contempt is usually said to involve a negative evaluation of other people and their actions (Rozin et al., 1999). The emotion is often linked to hierarchy and a vertical dimension of social evaluation (Miller, 1997).

Even though contempt and disgust overlap in some ways, there are some important distinctions. Miller (1997) argues that both feelings assert a superior ranking against their objects but the experience of superiority differs. While feelings of contempt can make us feel pride and superiority over others, disgust makes us pay with sensations of unpleasant sensations for the superiority it asserts. There is also evidence that contempt and disgust have different neural substrates. Sambataro et al. (2006) used fMRI to investigate the neural basis of contempt and disgust and found some significant differences. While contempt was associate with activity in the amygdala and the globus pallidus and putamen, disgust showed greater activation in the right insula and caudate. The authors concluded that the results indicated involvement of different neural substrates in the processing of facial emotional expressions of contempt and disgust.

Haidt (2003) argues that a prototypical moral emotion has two features: disinterested elicitors and pro-social action tendencies. If contempt can lead to feelings of pride and superiority, this emotion could have strong and direct benefits to the self, thereby making its elicitors more egoistic than disinterested. Maybe people actively seek objects to direct their contempt towards in order to feel pride and superiority. One could of course make the argument that all emotions have developed in order to indirectly benefit the self (Haidt, 2003). However, contempt seems to more egoistically than altruistically motivate judgement and behaviour, and therefor its place among the moral emotions might be questioned.

However, Miller (1997) argues that contempt also has a softer side. The notion of looking down at people, he claims, can sometimes awake softer and gentler sentiments like pity, love and graciousness (Miller, 1997). If this were the case, it would contrast disgust since disgust is linked to a reduced inclination to emphasise with the disgusting object. On the other hand, Haidt (2003) claims that contempt also can weaken other moral emotions such as compassion, making us colder towards the victimized individuals. If contempt would have a strong link to compassion, it might be more qualified as a moral emotion. Evidence of this, however, seems insufficient.

In *moral judgement* and *moral behaviour*, Hutcherson & Gross (2011) argues that contempt can be seen as an emotion strongly concerned with appraisals of incompetence. In our concern with people's competence, contempt might work as a force that helps diminish interactions with lower rank individuals who cannot make meaningful contributions to the group. Further, Miller (1997) argues that contempt also is important in maintaining social structures and hierarchies. Even though people in different cultures and societies put different emphasis on authority and social hierarchies, Haidt (2012) argues that authority and hierarchical structures, if not misused, help protect order and fend of chaos.

Although contempt might work as a force that maintains hierarchical structures by making people sensitive to disobedience, disrespect or rebellion, the underlying motives of contempt might be driven by selfishness and an inclination to look down on others in order to experience superiority and pride. The social benefits of contempt might therefore be nothing but a bi-product of selfish drives. Its place among the moral emotions could therefor be questioned, especially if we define moral emotions as those emotions that altruistically motivate behaviour.

The self-conscious emotions.

Shame, guilt, embarrassment and pride are all members of the family of self-conscious emotions because they are evoked by self-reflection and self-evaluation. This self-evaluation can be implicit or explicit but most important is that the self is the object for these emotions (Tangney et al., 2007). However, there are some important differences. Evidence suggests that these are distinct emotions with different phenomenological features (Tangney & Miller, 1996). They also seem to be triggered by different situations and vary in their action tendencies (Tangney & Miller, 1996). The self-conscious emotions together appear to function as an emotional moral barometer, providing immediate and salient feedback on our social and moral acceptability (Tangney et al., 2007). Haidt (2003) argues that the self-conscious emotions seem designed to help people fitting into groups without triggering contempt, anger and disgust in others. They do this by providing immediate feelings of reinforcement (pride and self-approval) when we “do the right thing”, or direct punishment (shame, guilt or embarrassment) when we “sin” (Tangney et al., 2007).

Embarrassment and shame. Although shame and embarrassment are closely connected there are some important differences in elicitors and action tendencies. The self-conscious emotions all have in common that they depend on our ability to understand and be concerned with what other people think of us (Miller, 1996). This ability is generally referred to as empathy. Baron-Cohen (2012) defines empathy as the ability to identify what someone else is thinking and feeling and respond to their thoughts and feelings with an appropriate emotion. Although there are no available brain imaging studies on shame, the neural correlates of guilt and embarrassment were examined in an fMRI study by Takahashi et al. (2004). The authors hypothesized to find similar activation pattern as in empathy-related areas, since both emotions depend on our ability to understand and be concerned with what other people think of us. The result showed that both guilt and embarrassment commonly activated the medial prefrontal cortex (MPFC), left posterior superior temporal sulcus (STS) and visual cortex. Compared to guilt, embarrassment produced greater activation in the right temporal cortex, bilateral hippocampus and visual cortex. Most of these regions have been identified in the neural substrates of theory of mind (ToM). These results might therefore support the idea that these emotions require the ability to represent the mental states of others (Takahashi et al., 2004). Even though the neural correlates of shame has not yet been examined, Moll et al. (2008) hypothesize that shame should activate similar brain areas as in embarrassment and guilt.

Embarrassment is often elicited in intrapersonal context and arises when we receive unwanted attention from others (Prinz, 2007). It is a negative feeling but it is often negative in a light-hearted way. Shame is more serious and arises when we have done something morally wrong (Prinz, 2007; Tangney et al., 2007). Tangney and Miller (1996) found that shame was experienced as a more intense and painful emotion that involved a greater sense of moral violation. Ashamed people were more angry and disgusted with themselves and also felt more regret and responsibility. In contrast, embarrassment arose from more trivial and humorous events and was accompanied with obvious physiological changes like blushing and increasing heart rate (Tangney & Miller, 1996). Embarrassment is often elicited in front of an audience while shame can be felt when people are alone (Tangney & Miller, 1996). Embarrassment can also be felt in the presence of higher-ranked individuals, especially in cultures with rigid systems of social stratification (Prinz, 2007: Haidt, 2003).

Some argue that embarrassment is a less moral emotion than shame because embarrassment often is elicited by violation to social conventions while shame relates more to violations of moral norms. Even though there might be some truth to this claim, they both still have impact in *moral behaviour* (Haidt, 2003). Because of their submissive nature (Haidt, 2003), they both promote defensiveness, interpersonal separation and distance (Tangney et al., 2007). They can also inhibit assertive behaviour, thereby reducing the likelihood of punishment from dominant others (Haidt, 2003). Because shame is a painful emotion, it is often believed that it will motivate people to avoid doing wrong because the anticipated fear of experience shame. However, evidence of this claim seems to be insufficient (Tangney et al., 2007). Rather, it appears that shame has a dark side in moral behaviour. Firstly, it is believed to decrease empathy and empathic responses toward others. Shame is often a painful, self-oriented distress that comes from the concern of other people's evaluation of us (Tangney et al., 2007). In the shame experience, there is no clear distinction between self and behaviour, and the malignancy of the behaviour is quickly generalized to the entire self. Since shame promotes a self-focused distress reaction, it can thereby make it difficult for people to focus and maintain other oriented empathic responses (Tangney, 1991). Secondly, shame has also shown to be correlated with anger. Feelings of anger can make shame-prone individuals more likely to engage in externalization of blame and express anger in destructive ways including physical, verbal and symbolic aggression (Tangney et al., 2007). Viewing shame in this light, one might question its place among the moral emotions. Although the elicitors might be disinterested, shame does not seem to have strong pro-social action tendencies. However, does shame always leads to external blame or can it also motivate people to right the wrong? If it can, under what circumstances? There seems to be two different elicitors of shame: one that has to do with one's self image and one that has to do with the concern for one's social reputation. Lets say I steal money from my best friend and he never finds out about it. I might still feel ashamed because I realize that I'm not a very good friend (self-image damaged). Or lets say he does find out about it and tells everyone that I am an unreliable and selfish friend. I might then feel ashamed because other people know what a horrible person I am (social reputation damaged). Do these two different elicitors have distinctive effects on pro-social behaviour? Are there any other factors that determine how one responds to shame? Even though shame has a darker side in moral behaviour, maybe one should not so easily dismiss it as a moral emotion without further investigate if personal and environmental factors determine different pro-social responses to shame.

Guilt. Guilt is sometimes confused with shame but Haidt (2003) argues that the two of them seems to grow out of different psychological systems. This is, however, yet to be confirmed with brain imaging studies. The MPFC, the STS and the anterior cingulate cortex have all shown to be involved in the experience of guilt (Takahashi et al, 2004). However, as previously noted, no brain imaging study on shame has been conducted.

Although shame and guilt both seem to be induced in public situations and equally linked to interpersonal concerns, there seems to be a systematic difference in the nature of those concerns (Tangney et al., 2007). One important difference is that in shame, the core self is at stake, while guilt is more often linked to a specific behaviour or act (Tangney et al., 2007). Tangney (1991) relates guilt to specific behaviours or acts that are somewhat apart from the self. This separation between the self and a specific act or behaviour takes individuals one step closer to the distressed other, and thereby closer to an empathic response (Tangney, 1991). There is in other words a difference in “egocentric” versus “other-oriented” concerns when it comes to shame and guilt (Tangney et al, 2007), making guilt a more adaptive emotion since it is more beneficial for individuals and relationships (Tangney, 1991).

In *moral behaviour*, guilt is invoked by people to apologize for misdeeds, to express sympathy, to discipline children, to increase self-control, to refuse sex and more. It also motivates people to perform or avoid a stunning variety of actions because of the anticipation of guilt (Baumeister et al., 1994). As mentioned before, guilt is also positively linked to empathy. Baumeister, Stillwell and Heatherton (1994) argues that since experience of guilt is specifically linked to the bad behaviour, it in turn highlights the negative consequences experienced by others. Thereby guilt fosters an empathic response and motivates people to “right the wrong”. This finding has been supported by later findings by Tangney, Stuewig, Mashek & Hastings (2011). In their study on 550 jail inmates, they found that guilt-prone inmates exhibited more empathy and lower levels of externalization of blame and hostility, relative to those who were less guilt-prone.

In sum, guilt seems to qualify as what Haidt (2003) would call a moral emotion since it both appears to have disinterested elicitors and pro-social action tendencies.

OFC-damage disrupts the self-conscious emotions. The self-conscious emotions require, as early noted, the ability to evaluate one's self and to infer the mental states of others. They are believed to function as reinforcement to socially appropriate behaviour, and, when expressed, also repair social relations following transgression (Beer Heerey, Keltner, Scabini & Knight, 2003). The frontal lobes have been characterized as centres of regulation or executive control, where the orbital regions of the frontal lobes seems particularly involved in the regulation of social behaviour (Beer et al., 2003). They are highly connected to emotional and social processing areas, such as the amygdala, and when damaged, seem to disrupt social regulation, decision-making and one's emotional life (Damasio, 1994; Bechara, Damasio & Damasio, A., 2003). Beer et al. (2003) found that impaired behavioural regulation of patients with OFC-damages was associated with disrupted self-conscious emotions. Although the patients did generate self-conscious emotions, they tended to reinforce rather than correct inappropriate behaviour. They could for example feel pride, rather than embarrassment, when engaging in improper behaviour. And if they felt embarrassed, it only tended to reinforce their belief that their social behaviour was exemplary (Beer et al., 2003). Additionally, Beer et al. (2003) also found that the OFC patients seemed to have a reduced ability to recognize other people's evaluations of their actions. An important part of recognizing other's feelings is to be able to read facial expressions. OFC patients had trouble recognizing other people's self-conscious emotions, something that could disrupt their own ability to feel self-conscious emotions since they where unable to benefit from other people's feedback (Beer et al., 2003). One should, however, be cautious about drawing conclusions about the specific role of the OFC. Other areas, such as the amygdala, are connected to the OFC through a neural circuitry and also have shown to impair recognition of other's emotional state (Aldolphs et al, 2005). Further studies will be needed in order to draw more specific conclusions about the exact role of the different brain areas discussed above.

The other-suffering emotions.

Empathy/sympathy. A basic fact about human beings is that we feel bad when other people suffer and sometimes we are so moved by these feelings that we decide to help (Haidt, 2003). There are many words for this ability including empathy, sympathy and/or compassion. It is important to note that empathy, sympathy and compassion are not discrete emotions with a distinct physiology or facial expression. Rather it is an emotional process with substantial implications for moral behaviour (Tangney et al., 2007). Current conceptualizations of empathy integrate both cognitive and affective components. The cognitive component is the identification of other people's thoughts, desires, beliefs, intentions and goals. Some refer to this ability as Theory of Mind (ToM) and it relies on structures of the temporal lobe and the pre-frontal cortex (Singer, 2006). The affective component refers to the ability to identify other people's feelings (emotions and sensations) and it relies on sensorimotor cortices as well as limbic and para-limbic structures (Singer, 2006). This section will focus on the affective component of empathy. However, it is important to note that in empathy, the ability to distinguish between self and other is crucial (Singer & Lamm, 2009). In other words, we must be able to distinguish between whether the source of our affective experience lies within ourselves or was triggered by someone else. Without this ability we would witness someone else's distress, believe it was our own, and have a self-centred response without any concern for the person who suffered (Singer & Lamm, 2009).

Empathy, compassion, empathic concern and sympathy all have in common that affective changes are induced in the observer in response to the perceived or imagined affective state of another person. However, while empathy involves shared or isomorphic feeling of another, Singer and Lamm, (2009) argue that compassion, empathic concern and sympathy do not necessary involve shared feelings. In empathy, watching another person sad will make the observer feel sadness while in sympathy, compassion or empathic concern, another person's sadness might evoke feelings of compassionate love or pity in the observer (Singer & Lamm, 2009).

The separation between "feeling with" and "feeling for" has impact for *moral behaviour* in several important ways. Compassion, empathic concern and sympathy are more "other oriented" emotions that more often lead to helping behaviour (Tangney et al., 2007). Importantly in theses cases is that the empathic individual's focus remains on the experiences and needs of the other person, not on his or her own empathic response. In contrast stands the self-oriented distress that has a primary focus on the feelings, experiences and needs of the empathizer (Tangney et al., 2007), leading the affected party to selfishly alleviated its own distress (de Waal, 2008).

Empathy modulation. Even though empathy and empathic concern play a big role in our social and moral lives, our empathic concern does not stretch towards all humans and animals. In the search for how our relationship to others affect how we respond to their emotions, Singer et al. (2006) found that empathy was modulated by the perceived fairness of others. In their study they used an economic game model to induce liking or disliking of two confederate actors, previously unknown to the experimental subjects. In the first part of the experiment the confederates played both fair and unfair. In the second part the experimental subjects observed when the fair and unfair confederates received painful electrical shocks. The fronto insular (FI) and anterior cingulate cortices (ACC) have shown to play a crucial role in experience one's own and other's pain. When the participants observed fair players in pain, both sexes exhibited empathy-related activation in these pain-related areas. However, in males these empathy-related responses were significantly reduced when observing an unfair person receiving pain, while women's empathic responses slightly decreased. The reduced empathy-response in men was accompanied by activation in reward-related areas, with an expressed desire for revenge. The authors concluded that the empathic responses were shaped by the evaluation of other people's social behaviour, such as an inclination to empathize with fair players and favouring physical punishment of unfair players (Singer et al, 2006).

The other-praising emotions.

The emotions so far have mostly been concerned with bad deeds made by others or by the self, or bad things experienced by others. The brighter side of the moral emotions, however, is that people also seem to be emotionally sensitive to the good deeds and moral examples of others (Haidt, 2003). Mainly two positive emotions seem to meet the demands of Haidt's (2003) definition of a prototypical moral emotion; disinterested elicitors and pro-social action tendencies. These two are gratitude and elevation and they are produced by the good or virtuous deeds of others.

Gratitude. Gratitude is a pleasant affective state that generally is felt in response to another person's benevolence (Tangney et al., 2007). Recent studies using game theoretic methods have started to investigate the brain regions involved in positive reciprocity, which has clear implications for the neural underpinnings of gratitude. Brain regions activated when people saw faces of good and reliable cooperation partners included the ventral striatum (including nucleus accumbens and putamen) and bilateral OFC (Singer et al, 2004). The two most robust predictors of gratitude are that the benefactor is being responsive to the recipient's needs and wishes, and that the receiver likes the benefit (Algoe & Haidt, 2009). McCullough, Kilpatrick, Emmons & Larsson (2001) argues that gratitude serves two functions: it is a response to moral behaviour and a motivator/reinforce of moral behaviour. The response function as an emotional moral barometer, reliably detecting that one has been benefited from the actions of another moral agent. Gratitude can also work to motivate and reinforce pro-social behaviour (McCullough et al., 2001). As a *motive*, people that have received help from a benefactor might be motivated by their gratitude to reciprocate. However, it is not clear if people's need to reciprocate comes from their feelings of gratitude or from the feelings of indebtedness (McCullough et al., 2001). Gratitude can also work to *reinforce* pro-social behaviour. However, the pro-social action tendencies seem to be dependent on whether one expresses gratitude to a benefactor or not. People who are thanked for their pro-social behaviour are more inclined to help their beneficiaries again. They are also more likely to help third parties after being shown gratitude by the initial beneficiary (McCullough et al., 2001). Seen from an evolutionary perspective, gratitude can be seen as a motivator of reciprocal altruism, encouraging people to repay benefactors, just like moral aggression motivate people to punish cheaters and free-riders (Travis, 1971).

Elevation. Elevation is characterised by Haidt (2003) as a prototypical moral emotion since it has very disinterested elicitors and high pro-social action tendencies. Elevation is elicited by acts of charity, gratitude, generosity, fidelity or any other strong display of virtue. It is a response to acts of moral beauty in which we feel less selfish and want to act accordingly (Algoe & Haidt, 2009). Elevation also have distinct physical feelings that manifests through a sensation of "opening in the chest" combined with the feeling that one has been uplifted or elevated in some way (Algoe & Haidt, 2009). The neural substrates of elevation and other emotions that are elicited by the virtuous and excellence of others have not received as much attention as those emotions that give rise to negative evaluations of others (Englander, Haidt & Morris, 2012). However, a recent study by Englander et al. (2012) found that the experience of elevation and admiration were associated with brain regions including the medial prefrontal cortex, precuneus and insula. These regions have previously been implicated in self-referential and interoceptive processes (Englander et al, 2012). In moral behaviour, elevation might motivate people to do charitable and grateful acts, but empirical work supporting this theory is inadequate (Algoe & Haidt, 2009).

In sum, this presentation of the neural basis of the moral emotions and their effects on moral judgement and moral behaviour is no way a complete or inclusive model. Although they appear to be linked to distinct neural systems in the brain and affect moral judgement and moral behaviour in different ways, further research should aim to specify these links.

Moral reasoning

Defining moral reasoning

Moral reasoning contrasts moral intuitions in several ways. While intuitions are fast, automatic and affective (Haidt, 2001), reasoning is slow, deliberate and conscious and requires attention and effort (Kahneman, 2011). We use both intuitions and reasoning in our daily lives, many times outside the moral sphere. A definition of moral reasoning is therefore needed in order to separate moral reasoning from other types of reasoning. Haidt (2001) defines moral reasoning as “conscious mental activity that consists of transforming given information about people in order to reach a moral judgement” (p. 818). However, Paxton & Greene (2010) argues that this definition is too broad, because it allows any conscious thought process about people that affects moral judgement to count as moral reasoning. For example, imagine a police officer that is about to determine whether or not a man is guilty of his wife’s death. Fingerprints and DNA evidence indicate that the man killed his wife and this evidence leads him to believe that the man is guilty. Should one claim that the police officer engaged in moral reasoning? Or imagine someone saying that “Anti-war protesters are communist, fascist, pigs who should go back to Russia!” (Paxton & Greene, 2010, p. 5) Is this moral reasoning? In both examples one uses conscious mental activity that consists of transforming given information about people in order to reach a moral judgement (Paxton & Greene, 2010).

Is one reason or justification enough to claim that we have engaged in moral reasoning, no matter how unjustified or inconsistent the reason is? According to Haidt (2001) it seemingly would be. However, Paxton & Greene takes a more narrow definition, claiming that moral reasoning is an attempt to compel oneself or another individual to accept a moral conclusion on pain of inconsistency. They define moral reasoning as “ conscious mental activity through which one evaluates a moral judgment for its (in)consistency with other moral commitments, where these commitments are to one or more moral principles and (in some cases) particular moral judgments" (p. 6). Therefore, in moral reasoning, a judgement is not merely altered through conscious mental activity, but it is altered in a *principled way*, so as to be consistent with one's other moral commitments (Paxton & Greene, 2010). Paxton & Greene (2010) seem to draw a distinction between "good reasoning" (logical consistency) and biased or unproductive reasoning, where the last mentioned would not qualify as moral reasoning. At first sight, a narrow definition might appear to be prudent from a theoretical point of view. It corresponds in many ways to Galotti's (1989) reflection of reasoning, where he claims that it could be defined as "thinking according to the theorems of a logical system" (p. 332). However, this definition immediately faces many problems (Galotti, 1989). If reasoning is defined narrowly as solving problems or making a judgement based on principles of logic and consistency, and Haidt (2001) is right in his claim that this form of reasoning in the moral sphere is rare, then why should we study it? Why build a theory on moral reasoning dependent on logical consistency when this is not something we generally do outside the laboratory (Galotti, 1989). At the same time, having broad definitions in a theory is not ideal since it might compromise the validity of general conclusions. One way to solve this problem could be to distinguish post hoc moral reasoning (moral rationalization) from more principled moral reasoning. Haidt (2001) argues that moral reasoning is mostly a post hoc affair; we decide what is right and wrong based on intuitions and then, if necessary, we come up with explanations that justifies or explains our judgements. However, in a study by Paxton & Greene (2010), they found that they could change people's judgement by instructing them to make more rational, objective judgement about moral dilemmas. So maybe people can change each other's judgement, not by modifying the target's intuition, but by appealing to their ability to reason, and to make judgements that are consistent with their other moral commitments (Paxton & Greene, 2010). This type of reasoning might be rare, as Haidt (2001) claims, but evidence suggest that it might occasionally happen. However, only because one can influence a person's judgement by appealing to their ability to reason, it does not follow that one actually changes the target's judgement. Maybe people are just telling what they think other people want to hear. Even though this might be the case, it is still interesting that we actually seem to be able to influence other people's judgement by appealing to their reason.

Greene (2007) argues that there are two ways of reasoning that are related to different psychological systems in the brain. The first one is more driven by emotional responses with reason as an agent engaging in post hoc moral rationalization. The other way of reasoning is more “cognitive” and more likely to involve genuine moral reasoning (Greene, 2007).

This distinction between moral reasoning and moral rationalization is crucial in order to understand why we reason and how moral reason relates to our intuitive system and to moral behaviour.

Intuition-based reasoning

Hume’s argument that “reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them” (Hume, 1969, p. 462) has gained support from recent studies in the neuroscience of morality and human decision-making (Damasio, 1994; Haidt, 2001; Greene, 2007). Hume saw our passion (emotions/desires) as the motive behind actions, while reason itself could never motivate or inhibit an action. If he is right, how can this be? Why should our adaptive moral behaviour be driven by moral emotions as opposed to moral reasoning?

Churchland (2011) argues that our emotional system govern our behaviour in several ways. Some emotions, including sensations of cold, warm, thirst, hunger and pain, govern the organism’s inner environment in order to ensure that the inner state of the body is within the parameters that matter for survival. The social emotions, including anger, shame, guilt, disgust, sympathy and pride, are nature’s way of guiding us through the social sphere. Our punish/reward system is a way of learning to use past experiences in order to improve our performance in both domains (Churchland, 2011). Imagine that Mother Nature instead would hand over this job to reason. Is it reasonable to believe that a candidate that is slow, deliberate, takes a lot of energy and can generally only focus on one thing at the time (Kahneman, 2011), is suitable for controlling our bodies inner environment and at the same time make fast and wise decisions in complex social situations? Emotions have been prioritized because they deliver quick, reliable and efficient responses to re-occurring situations and events, whereas reasoning is unreliable, slow and inefficient in such contexts (Greene, 2007). Damasio (1994) provides a good example to what might happen if we don’t have access to automatic decision-making abilities, while our abilities to reason are somewhat intact. This peculiar combination is sometimes manifested in people with damages to the ventromedial region of the PFC. One day, Damasio was discussing two alternative dates for a new appointment with one of his patients with this type of brain damage. The patient started to provide reasons for and against the two proposed dates for over half an hour. He mentioned other commitment that might collide with the two dates, considered possible weather conditions and many other things most people would never think about. After a while, when he was lost in never ending cost-benefit analyses, Damasio interrupted the patient and said that he should come the latter of the two dates. The patient calmly agreed and left the room.

This is a good example of the limitations of reasoning and how devastating it might be not to have access to automatic decision-making processes (Damasio, 1994). The faculty of reasoning in the brain simply does not work effectively without emotional reactions, which might suggest that biological drives, body states and emotions may be and indispensible foundation for rationality (Damasio, 1994). The lower levels in the neural structures of reason are the same that regulate the processing of emotions and feelings, along with global bodily functions such that the organism can survive. Rationality, therefor, is probably shaped and modulated by body signals, even as it preforms the most subtile distinctions and act accordingly (Damasio, 1994). Although many important choices involve feelings, there are also many choices that we make without a consciously, distinct experienced emotion. However, this does not mean that the dispositional machinery underlying the process has not been activated (Damasio, 1994). If we are in a situation where we have to make a choice, possibilities must be ranked and placed in an order. If they are ranked, then criteria are needed (values or preferences), which are given by our affective system. This system might affect how the brain handles combinations of images, and thus operate as a bias. This bias could affect where we direct our attention and how we choose to allocate attentional enhancement differently to each component (Damasio, 1994). In other words, according to Damasio (1994) our affective system affects our rational considerations in many ways. It selects available options, ranks them and guides our attention through better or worse alternatives. Mercier and Sperber (2011) support Damasio's view about how our intuitive system affects the way we reason. They claim our intuitive system and many other networks in the brain operate mainly outside the realm of awareness, each providing specialized bits of information (Mercier & Sperber, 2011). Even though people can be aware of having reached a certain conclusion, we are sometimes only aware of the result of this conclusion and not of the many intuitive processes behind it (Mercier & Sperber, 2011). This form of judgement is called an intuitive belief: a belief that has been reached by an unconscious process that makes us hold a belief without awareness of reasons to hold them (Mercier & Sperber, 2011). A reflective belief on the contrary, is a belief that is held with awareness of one's reasons to hold them. Mercier & Sperber (2011) do not deny that one may arrive at a belief through reflecting on reasons to accept it. They call this form of reasoning as *reason proper*, which is characterized by the awareness of a conclusion and the argument that justifies accepting that conclusion. However, they suggest that the arguments exploited in reasoning are the result of an intuitive inferential mechanism that is unconscious and intuitive. In other words, arguments are not the result of an explicit reasoning system that would stand apart from the intuitive system. Rather, arguments are the result of one mechanism of intuitive inference among many that delivers intuitions about premise-conclusion relationships. We use our intuitive system to evaluate arguments either as strong or as weak. These evaluations and preferences are ultimately grounded in intuition (Mercier & Sperber, 2011).

If intuitions are deeply imbedded in our reasoning system and affect the way we think and act, then one should be able to affect people's judgements by changing their intuitions. Greene's et al. (2001) fMRI study on people's judgements concerning personal and impersonal moral dilemma provide support for this claim. A typical impersonal moral dilemma is a variant of the trolley problem. In this dilemma, a run-away trolley is heading down towards five people who are working on the track. They will be killed if the trolley proceeds on its course. The only way to save them is to hit a switch, which will turn the trolley away from the five workers and on to a side-track where it will kill one person. In this scenario most people say that it is ok to turn the trolley onto the side-track where it would kill one person instead of five (Greene et al., 2001). The reasons are more utilitarian in this case since people argue that five lives are better than one. However, in what Greene (et al., 2001) calls a personal dilemma, a run-away trolley is heading down the track where five people are working. You happen to stand on a bridge above the track and next to you stands a large man, and you can save the five people by pushing the man off the bridge onto the tracks. However, in this case, the majority of people disagree of killing one person in order to save one (Greene et al., 2001). Greene et al. (2001) believes that the emotional response is likely to be the crucial differences between the two cases. The direct physical harm that we exert in the personal dilemma in order to push the large man over the bridge seem to elicit stronger negative emotions than pulling a switch. In support for this claim, Greene et al. (2001) found that areas that exhibited increased activity when participants considered personal moral dilemmas were those associated with emotion and social cognition (medial prefrontal cortex, posterior cingulate/precuneus, and superior temporal sulcus /temperoparietal junction).

Characteristically reasons for not pushing the fat man onto the tracks vary. Some will be tautological: “Because it’s murder”! Others are more sophisticated: “The ends don’t justify the means”. “You have to respect people’s rights” (Greene, 2007). However, these arguments could as well be used in the first case where one can hit a switch and turn away the trolley onto the side-track where it would kill one person. The talk about rights and respect for individuals are only natural attempts to rationalize what we feel when we find ourselves having strong emotionally-driven intuitions that are in odds with more “cold” calculus of utilitarianism (Greene, 2007).

Another example of a personal dilemma is Peter Singer’s (1972) scenario about a drowning child. Imagine that you are out walking and happen to see a child drowning in a shallow pound. Most of us feel a very strong obligation to help the child, even if it means getting one’s clothes dirty. However, this principle is not something we generally do not apply when it comes to donating money to far away starving and dying children (Singer, 1972). Instead we spend a great deal of our money on unnecessary luxuries. One might rationalize this behaviour by claiming that aid is mostly ineffective, only serving enrich corrupted politicians or creating more poor people (Greene, 2007). Many normative explanations might come in mind, but non seem very compelling. Yet people tend to believe that they are within their moral right to spend money on luxury for themselves, even though the money could drastically improve the lives of many others. This is exactly what one would expect if our sense of moral obligation were driven by emotions and if emotions tended to be most strongly triggered by personal events (Greene, 2007).

A third example of how our emotions shape our judgement, reason and behaviour comes from studies on psychopaths. Psychopaths appear to have no cognitive deficit in understanding other’s state of mind, including their beliefs, desires, motives and intentions (McGeer, 2008). However, in contrast to this cognitive capacity they have been found to be notably abnormal in their affective profile, seemingly insensitive to the suffering and distress of others (Haidt, 2001). Blair (1995) argues that this might explain why psychopaths have difficulties to distinguish between moral and conventional transgressions. Although the difference between moral and conventional transgressions might not be clear-cut, one difference is that moral transgressions are often considered more seriously wrong than conventional transgressions because they provoke a strong affective response in us (Haidt, Koller & Dias, 1993). Thus, for instance, we code those transgressions that result in physical or psychological suffering of victims as paradigm moral transgressions because of our affective response to the victim's imagined distress, something to which the psychopath is apparently blind (McGeer, 2008). Hence, the psychopath fails to distinguish more "severe" moral transgressions from those that merely break the accepted rules of social life. This seem to support the view that our capacity for moral thought and action is heavily dependent on our affective nature and to the capacity to empathically respond to the affective state of others (McGeer, 2008).

In sum, both our moral behaviour and our way of reasoning seem heavily dependent on our emotional reactions. This clearly supports Haidt's (2001) view that moral reasoning might often be a post hoc rationalization of our intuitions. If it is the case that our judgement and behaviour is primarily driven by intuition and not by reason, then what is the function of moral reasoning?

Reputation and accountability.

Human beings are extraordinary co-operators, and we live in a world were cooperation beyond kinship is extensive (Haidt, 2013). Cooperation beyond kinship is usually explained in terms of reciprocal altruism: you scratch my back and I'll scratch yours (Singer, 2011).

The largest threat to reciprocal altruism is individuals who gladly take but never reciprocate. This problem might on one level be solved by abilities to recognize and remember who is a good co-operator and who is not, and to punish individuals who cheat (Singer, 2011). In this way cheats never prosper because their selfishness is noticed and punished. Reciprocal altruism is also more likely to be found in species with a relatively long life-span, living in small stable groups (Singer, 2011). The increase of group size is valuable when groups are defending themselves towards predators but it also demands more complex cognitive abilities in order to handle the social cognitive demands of managing proportionally more relationships (Dunbar, 2004). Speech and the exchange of information provide a huge advantage because it allows humans to interact with a larger number of individuals simultaneously and exchange information about the state of our social network (Dunbar, 2004). Language serve many important functions, one of them is reputation. Not only can we identify free-riders, but we can also tell other people about their selfish inclination, thereby reducing other people's willingness to cooperate with them in the future (Singer, 2011).

The fear of bad reputation leads people to be what Tetlock (2002) calls "intuitive politicians". As object of accountability pressures from others, people strive to maintain positive social identities towards significant constituencies in their lives. He further argues that people at the same time are like intuitive prosecutors who try to detect cheaters and free-riders that seek the benefit of exploiting the group. A key function of thought is to close loopholes in accountability regimes that free-riders might otherwise exploit (Tatlock, 2002). As prosecutors, we are like scientists trying to figure out the motive behind people's actions. The motive others attribute to someone's action forms that person's reputation and affects to what extent people will hold the person accountable for the action (Tatlock, 2002). When we have to justify our actions or beliefs to other people, we normally engage in two different types of reasoning: pre-emptive self-criticism and defensive bolstering (Tatlock, 2002). Pre-emptive self-criticism is the strategy of anticipation plausible objections from critics, factoring those objections into one's mental representation of the problem in order to reach a complex synthesis that specifies how to deal with trade-offs. This seems to be a more top-down way of reasoning, which might be closer to what Paxton & Greene (2010) would call "genuine" moral reasoning. However, people only engage in this complex, energy-consuming and self-critical way of reasoning when the situation strongly motivates us to do so. Lerner & Tatlock (1999) found that pre-emptive self-criticism is especially likely when people are unconstrained by past commitments, the views of the audience are unknown or known to conflict, the audience is knowledgeable and powerful, and when the audience possesses a legitimate right to question the reasons behind opinions or decisions. Defensive bolstering on the other hand contrasts pre-emptive self-criticism. While accountability still motivates thought, it takes a self-justifying rather than a self-critical thought. Here people devote mental energy in order to generate reasons why they are right and their critics are wrong. This form of reasoning is more likely when they are held accountable for something they have undeniably done, but have the chance to present their story in a more positive light in order to hide weak competence or morals, thereby protecting their reputation (Tetlock, Skita & Boettger, 1989). Defensive bolstering (Tatlock, 2002) or moral rationalization (Greene, 2007: Haidt, 2001) seems to have several things in common. They are all a biased way of reasoning and are often triggered in intrapersonal contexts where people feel pressure from others to provide reasons for actions, thoughts or beliefs. People might be able to exert more “objective” or “genuine” self-critical reasoning but only under very specific circumstances where we are unconstrained by past commitments and where the audience is knowledgeable and powerful (Tatlock, 2002). It appears strange, in a world of accountability pressures from others, that the only thing humans value when it comes to our own actions and beliefs is “truth”, no matter the consequences. As Tetlock (2002) writes: “People do not care about accuracy per se; they care only about justifiability, and justifiability is a profoundly relational construct that hinges on the identity of the audience and its evaluative standards” (p. 468).

Mercier & Sperber (2011) support this view. They argue that the main function of reason is to present evaluative and convincing arguments for their decisions. Because of this function, reason is often biased and drives people towards decisions that they can justify, even if these decisions are not optimal. Even though we might be able to reason “better” or “unbiased” in some cases, this form of reason seems rare and heavily dependent on Tetlock’s (2002) four criteria of intrapersonal contexts.

The interpreter.

If intuitions guide our thought, beliefs and actions in ways that are not always accessible in consciousness (Haidt, 2001: Damasio, 1994: Mercier & Sperber, 2011) it is questionable whether we even have the possibility of providing accurate explanations to other people about our motives behind attitudes, beliefs and actions. Yet we tend to experience and present ourselves as unified, conscious agents that appear to know exactly why we do the things we do and why we believe the things we believe? What part of the brain is responsible for these processes? Over the last 40 years, studies of split-brain patients have provided numerous insights into the processes of perception, attention, memory, language, reason and consciousness (Gazzaniga, 2000). In split-brain patients, the part of the brain (corpus callosum) that connects the right and the left hemisphere is cut of, which results in a loss of information exchange between the hemispheres. In the search for specific functions in the left and the right hemisphere, Gazzaniga (2000) found that the left hemisphere is, among other things, specialized in language and problem-solving. One of the major differences between the right and the left hemisphere is that the left appears to have a module that constantly forms hypothesis about events in the world and it looks for patterns even in the face of evidence that no pattern exists (Gazzaniga, 2000: Kahneman, 2011). Gazzaniga (2000) calls this module the interpreter, and the job of the interpreter is to make sense of all the internal and external stimuli that constantly is bombarding the brain. It does this by looking for relationships between events, searching for cause and effect, creating hypothesis and structuring a running narrative. The interpreter is the module that explains why we do what we do and why we feel what we feel (Gazzaniga, 2000). However, the interpreter can only use the information it receives. In split-brain patients, the left hemisphere has no access to what is going on in the right hemisphere. In an experiment where a negative mood was induced by visual stimuli in a patient's silent right hemisphere, the left hemisphere felt angry without having any idea why. However, when they asked the left hemisphere why it was upset, it immediately constructed a story and claimed that the experimenter was upsetting it. Even though the left hemisphere did not know where the negative mood came from, it immediately constructed a theory that explained its negative emotion (Gazzaniga, 2000).

The fact that the interpreter in split-brain patients constantly makes up stories and explanations to justify or explain choices and behaviours generated by the right hemisphere have been observed in several studies. However, one could argue that this form of post hoc explanation or justification is something normal people do, although perhaps not as obvious as in split-brain patients. We respond to the conscious deliverances of our unconscious perceptual, mnemonic, and emotional processes by fashioning them into a rationally sensible narrative, this without any awareness that we are doing so (Greene, 2007). In 1962, Schachter and Singer conducted a study where the participants were injected with epinephrine that activated the sympathetic nervous system. The symptoms are increased heart rate, hand tremors and facial flushing. The subjects were then put in contact with a confederate who behaved in either a euphoric or an angry manner. The participants who were informed about the effects of the drug attributed the physiological arousal to the drug. However, the participants who were not informed attributed the physiological arousal to the behaviour of the confederate. This is a good example of the human tendency to generate explanations for events. When we are aroused, we are driven to explain why. And if there is no obvious explanation, we generate one (Gazzaniga, 2000).

Genuine moral reasoning

While there is much evidence of the influence of automatic intuitive responses on moral judgement and moral reasoning, the role of reflective and self-critical reasoning remains uncertain (Paxton, Ungar & Greene, 2011). Paxton & Greene (2010) try to distinguish "genuine" moral reasoning from moral rationalization by defining moral reasoning as a conscious process where we evaluate a moral judgement for its (in)consistency with other moral commitments, such as a moral principle or a moral judgement. When we engage in this form of moral reasoning, we do not merely reach a judgement through conscious mental activity, but we reach it in a principled way, so as to be consistent with our other moral commitments (Paxon & Greene, 2010). This form of reasoning is not only distinct by its tendency to generate utilitarian moral judgement, but it also appears to be processed by specific areas in the brain (Paxton & Greene, 2010) including the DLPFC and the bilateral inferior parietal lobe (Greene et al, 2004). The anterior DLPFC is associated with abstract reasoning and cognitive control, and appears to compete with social-emotional responses to difficult personal moral dilemmas (Greene et al, 2004). One of these difficult moral dilemmas is called the crying baby dilemma (Greene, 2004). Consider the following scenario:

> It's wartime. You and your fellow villagers are hiding from nearby enemy soldiers in a basement. Your baby starts to cry, and you cover your baby's mouth to block the sound. If you remove your hand, your baby will cry loudly, and the soldiers will hear. They will find you, your baby, and the others, and they will kill all of you. If you do not remove your hand, your baby will smother to death. Is it morally acceptable to smother your baby to death in order to save yourself and the other villagers?

Many people find these types of moral dilemmas difficult, as indicated by relatively long reaction times (RT) and divergent judgement between subjects (Paxton & Greene, 2010). Greene (et al, 2004) found evidence for competing responses between emotional and "cognitive" brain areas, where more utilitarian judgements (smothering the baby in order to save yourself and other) were correlated with increased activity in parts of the DLPFC. Several other studies (Bartels, 2008: Moore, Clark & Kane 2008) supports that people with a greater working memory and more "rational" intellectual styles tend to give utilitarian answers to moral judgements. In sum, utilitarian judgements appear to be supported by moral reasoning because cognitive control requires the “top down” application of a guiding moral rule or principle (Paxton & Greene, 2010). To apply to a rule that overrides an intuitive response, one must determine that the intuitive response is incompatible with the rule, that is, engage in conscious moral reasoning. Paxton & Greene (2010) argue that in their research experience, subjects who make utilitarian judgements in response to moral dilemmas invariably justify their answers by appealing to utilitarian principles. There is also a body of evidence supporting that people with emotional deficits tend to make more utilitarian moral judgement. A study by Koenigs et al. (2007) tested people with focal damage to the VMPFC on personal and impersonal moral dilemmas. These patients are known for their defects in emotional responses and emotion regulation while the capacity for intelligence, reasoning and declarative knowledge of social rules is preserved (Damasio, 1994). In the impersonal moral dilemmas, Koenigs et al (2007) found no difference in judgement between VMPFC-patients and the control group. They all tended to give utilitarian answers to these dilemmas. However, in the personal dilemmas, there is typically a competing considerations of aggregate welfare on the one hand, and, on the other hand, harm to others that normally evoke a strong social emotion. Here the difference between VMPFC-patients and the control group was distinct. The VMPFC-patients had an abnormal high rate of utilitarian judgement, which Koenigs et al. (2007) attributed to their inability to experience pro-social emotions. These results do in some ways support Greene’s (et al, 2004) dual process theory in that moral reasoning (utilitarian cost-benefit analysis) is subserved by the DLPFC and tend to prevail when there is no prepotent emotional response. However, this interpretation faces some difficulties. Results from another study by Koenigs & Tranel (2007) challenge Greene’s (et al, 2004) hypothesis. They conducted a study on patients with damages to the VMPFC where they let the participants play a game called the Ultimatum game. The participants had to choose between accepting an unfair but financially rewarding proposal (an economically “rational” choice), or rejecting it to punish the unfair player (an ‘emotional’ choice). Here VMPFC-patients were more prone than controls to reject the unfair offer in order to punish the unfair player (i.e. they were more emotional than utilitarian). Greene’s dual process theory of competing systems of cognition and emotion is not necessary incompatible with these results. However, Moll and de Oliveira-Souza (2007) argue that that the dual process theory is not the only theory that can explain this effect. Instead of viewing emotion and cognition as competitive system, they argue that complex feelings such as compassion and other pro-social emotions emerge from integration between emotional and cognitive areas in the brain. While the more pro-social emotions are mediated by the VMPFC, other more self-centred emotions like anger and frustration are dependent on the lateral sectors of the DLPFC (Moll & de Oliveira-Souza, 2007). This might explain why VMPFC-patients are more “utilitarian” in personal moral dilemmas since they lack pro-social sentiments (Koenigs et al, 2007) and more “irrational” in the Ultimatum game (Koenigs & Tranel, 2007). In other words, Moll and Oliveira Souza (2007) emphasizes the degree of integration between the two systems as predictive of a judgement in contrast to Greene’s (et al, 2004) dual process theory of competing systems. A recent study by Tassy (et al, 2012) supports Moll and Oliveira-Souza’s integration theory. Tassy (et al, 2012) presented the participants with moral dilemmas while disrupting the right DLPFC through electric stimulation (TMS). The result were a higher level of utilitarian judgements, which does not fit with the dual system hypothesis predicting that right DLPFC should code for ‘rational’ cognitive control over emotional impulses. Instead the authors concluded that the right DLPFC might participate in the integration of representational emotions during moral evaluation, and therefor might not be responsible for the utilitarian bias.

Much is still to be said about the exact role of the DLPFC and its ability of principled and more “top-down” moral reasoning. The two theories discussed here are the integrative theory by Moll & de Oliveira Souza (2007) and the dual process theory by Greene (et al, 2004). In order for Greene’s dualistic theory to work, one would have to find double dissociation between utilitarian VMPFC-patients and anti-utilitarian DLPCF-patients. However, the evidence supporting VMPFC-patients tendency for utilitarian judgement appears stronger than the evidence supporting DLPFC-patients anti-utilitarian bias. Further research would need to specify the exact role of the DLPFC and how it relates to “emotional” parts in the brain.

Discussion

The aim of this essay is to :

(1) *Investigate the neural basis of moral intuition and moral reasoning, and see how the two systems relate to moral judgement and moral behaviour.*
(2) *Examine how the moral intuitive system and the moral reasoning system relate to each other.*

The neural basis of the intuitive system and how it relates to moral judgement and moral behaviour

In the beginning, I argued that morality has emerged from a coevolution of psychological mechanisms and cultural innovations in order for selfish individuals to profit from cooperation (Greene, 2013: Churchland, 2011: Haidt, 2008). The evolved psychological adaptations have enabled people to put collective interests before pure selfish interests, which is crucial in order for cooperation to thrive (Greene, 2013). The intuitive system appears to play a fundamental role in motivating both kin altruism and reciprocal altruism. Moral intuitions are distinguished from other intuitions by their tendency to altruistically motivate behaviour (Moll et al., 2008). Among these intuitions, we find the moral emotions that typically include shame, embarrassment, guilt, anger, contempt, disgust, empathy/sympathy, awe and elevation (Haidt, 2003). Empathy, which relies on sensorimotor cortices, limbic structures and structures of the temporal lobe and the pre-frontal cortex (Singer, 2006), enable us to care for other people's well-being in the same way we care for our own (Churchland, 2011). Empathy should not be viewed as an emotion. Instead, it is an ability to feel what someone else is feeling, including happiness, fear or boredom (Haidt, 2003). The effects of empathy on moral behaviour are dependent on whether we "feel for" or "feel with" the victim. While empathy involves shared or isomorphic feeling of another, compassion, empathic concern and sympathy do not necessary involve shared feelings. In empathy, watching another person sad will make the observer feel sadness while in sympathy, compassion or empathic concern, another person's sadness might evoke feelings of compassionate love or pity in the observer (Singer & Lamm, 2009). The more we can focus on other people's distress and not so much on our own empathic response, the more likely we are to help (Tangney et al., 2007).

Empathy also appears to play a crucial part in the experience of self-conscious emotions: shame, embarrassment and guilt. This is because they depend on our ability to understand and be concerned with what other people think of us (Miller, 1996). The orbital regions of the frontal lobes seem to be important for the ability to experience self-conscious emotions (Beer, 2003). Damages to this part have shown to impair people's ability to experience self-conscious emotions and often leads to a disruption in social regulation, decision-making and one's emotional life (Damasio, 1994; Bechara, Damasio & Damasio, A., 2003).

Among the self-conscious emotions, guilt has shown to be most positively linked to empathy (Baumeister et al., 1994) and tends to foster empathic responses and motivate people to "right the wrong". Shame on the other hand, has shown to have a darker side in moral behaviour. It has been correlated with anger, which can lead to physical, verbal and symbolic aggression (Tangney et al, 2007). Together the self-conscious emotions also function as an emotional barometer that provides feedback to our social and moral acceptability (Tangney, Stuewig & Mashek, 2007). They help us navigate through complex social context in order to fit into groups without triggering anger, contempt and disgust in others (Haidt, 2003).

Even the experience of moral anger has been connected to empathy (Batson, Shaow & Givens 2009). Moral anger has been viewed as an emotion triggered by a violation of a moral standard, but Batson, Shaw & Givens (2009) found that what makes people angry is maybe not the violation itself but the harm it does to significant others. Our judgements tend to be more severe when something wrong is done to people we care about, and our choice to act morally, or pro-social, might be determined by to what extent our action helps us to reach egoistic or altruistic goals (Batson, Shaow & Givens 2009).

While some of our emotions motivate us to supress selfishness by caring for other people and what they think of us, there are some emotions that serve to uphold reciprocal altruism and social structures. Free-riding is one of the largest threats to successful cooperation and nature seems to have solved this problem by developing emotions that make us sensitive towards people who do not contribute to the group. Haidt (2003) call these emotions the other-condemning emotions and they help us identify cheaters, liars, hypocrites and other who try to fake the appearance of reliable cooperation partners (Haidt, 2003). Anger, contempt and disgust are all a part of this family and are often triggered by violations to moral standards.

Moral disgust, which has been connected to activation in the orbital frontal cortex (OFC) and the insula (Moll et al, 2005a), is often elicited by moral violations like cannibalism, paedophilia, incest, betrayal and deception (Russell & Rogers, 2013: Haidt et al., 1997). Studies have shown that induces disgust tend to produce more severe moral judgements (Haidt, 2005; Shnall et al., 2008). Feelings of disgust can also lead to a reduced motivation to mentalize with the disgusting object, which can lead to dehumanization processes where we treat people as less human, hence excluding them from our moral concern (Sherman & Haidt, 2011). Contempt, which has been connected to activation in the amygdala and the globus pallidus and putamen (Sambataro et al, 2006), is an emotion strongly concerned with appraisals of incompetence (Hutcherson & Gross, 2011). It has suggested that it might work as a force that helps diminish interaction with people that cannot contribute to the group (Hutcherson & Gross, 2011). Further it also might be important for maintaining social structures and hierarchies (Miller, 1997).

Gratitude and elevation are two positive emotions that are elicited by the good or virtuous deeds of others (Haidt, 2003). The neural underpinnings of gratitude are not fully understood. However, studies on positive reciprocity has shown that brain regions activated when people saw faces of good and reliable cooperation partners included the ventral striatum (including nucleus accumbens and putamen) and bilateral OFC (Singer et al, 2004). Elevation has been associated with brain regions including the medial prefrontal cortex, precuneus and insula (Englander et al, 2012). These positive emotions can function to reinforce pro-social behaviour (McCullough et al., 2001), and they can also function as moral barometers that help us detect when we have benefitted from someone else. However, to what extent these positive emotions motivate pro-social behaviour is inadequate (Algoe & Haidt, 2009).

Although the moral emotions are just a small part of the intuitive system, they are important because they appear to be associated with certain parts of the brain, elicited by different events, form certain types of judgements and have a tendency to motivate altruistic behaviour. Since many of our judgements in the moral sphere appears to be the results of these emotions, understanding them might help us understand why we think the way we think and act the way we act. What is interesting about these emotions is that they appear to have been shaped by evolution in order to enable cooperation, but only with some people (Greene, 2013). Our moral intuitions, including the moral emotions, enable us to put “us” ahead of “me” but they have not evolved to put “us” ahead of “them”. We appear to be designed for cooperation, but only with some people. We don’t have a universal cooperation, because it would be incompatible with natural selection (Greene, 2013). Natural selection is a competitive process. The faster a lion can catch a prey, the greater advantage it has over other lions, and hence a better chance of reproduction. Natural selection would never work if there were no competition of resources (Greene, 2013). We appear to be the result of a long chain of natural selection where our moral intuitions have evolved because they gave our ancestor a greater chance to survive than other "less" cooperative groups. Morality can therefore be seen as a device evolved for successful intergroup competition. As a result, our moral psychology that enables successful cooperation within our own group undermines the cooperation between groups (Greene, 2013). If our judgement would primarily be shaped by moral intuitions that are biased towards certain people and situations, then one would expect to find large inconsistencies in people's moral judgements and behaviour. And this is exactly what we find. Take the example with the different versions of the trolley dilemma. Sometimes people are willing to take an action that will kill one person in order to save five, such as in Greene's (et al, 2001) impersonal dilemmas. However, as soon as you change the settings and make the dilemma more personal (trigger stronger intuitions), for example by having you push a guy onto the tracks in order to stop the trolley, suddenly most people judge the action to be wrong (Greene et al, 2001). As our emotional responses changes, our judgements tend to change. Or take the example with the study conducted by Batson, Shaw & Givens (2009) where participants listened to two stories about torture. In the first story, a marine from the USA, with whom the participant's shared national identity, was tortured. In the second story a Sri Lankan soldier, with whom the participants didn't share a national identity, was tortured. The violation is the same, and yet the first story produced much more anger than in the first story. The closer we are to the victim, the stronger emotion appears to be produced, hence a harsher judgement towards violation of moral principles.

We also seem sensitive to when others are harmed and we many times condemn these actions. However, as soon as we find out that the ones that are harmed are cheaters, violators or poor cooperation partners, our empathy for them drastically drops (Singer et al, 2006). Especially men can even experience pleasure when seeing a poor cooperation-partner receive electric shocks, with an expressed desire for revenge (Singer et al, 2006).

Maybe one of the best examples is Peter Singer's (1972) scenario about a drowning child. If we would be out walking and see a child drowning in a shallow pound, most of us would probably feel a strong obligation to help, even if it means getting our clothes dirty. If someone said that they didn't want to save the child because they wore new shoes or a nice suit, we would probably think he/she was a moral monster. However, this principle is not something we generally do not apply when it comes to donating money to far away starving and dying children (Singer, 1972). Instead we spend a great deal of our money on unnecessary luxuries.

If unbiased and consistent moral reasoning would drive our judgement and behaviour, these inconsistencies would be extremely hard to explain. Instead, it seems like one could make a strong case that our intuitive system strongly shape our judgement and behaviour, maybe to a greater extent than we would like to think.

The neural basis of the reasoning system and its affect on moral judgement and moral behaviour

In order to understand the moral reasoning system, one must understand its relationship with the intuitive system. People appear to engage in two different kinds of moral reasoning that relate differently to the intuitive system. The first one is more driven by emotional responses, with reason as an agent engaged in post hoc moral rationalization. The other kind of reasoning is more a top-down process where we form a judgement based on principles of logic and consistency (Paxton & Greene, 2010). Moral rationalization appears to be much more common than unbiased, logical and principled moral reasoning (Haidt, 2001; Paxton & Greene, 2010; Tatlock, 2002; Mercier & Sperber, 2011). The study of moral rationalization is important partly because it can provide an important insight concerning how and why humans reason, and partly because it can reveal how the reasoning system and the intuitive system commonly relate to each other.

The process of moral rationalization appears to be divided into two parts. The first step involves the production of a judgement based on intuition. Brain regions activated when people consider highly emotional moral dilemmas are those associated with emotion and social cognition, such as the MPFC, posterior cingulate/precuneus and superior temporal sulcus/temperoperietal junction (Greene et al, 2001). The second step is to rationalize the intuitive-based judgement, which appears to be done by a module in the left hemisphere called the interpreter (Funk & Gazzaniga, 2009). The interpreter is a pattern-seeking module that constantly comes up with explanations about why we do what we do and why we feel what we feel. However, the function of the interpreter might not primarily be to accurately describe “truth”. Instead, one of its important functions is to provide justifiable explanations to an evaluative audience in order to protect one’s reputation. Since free-riding and selfishness is one of the biggest threats to successful cooperation, our psychology is sensitive towards people who do not make valuable contributions to the group (Greene, 2013). This makes us what Tetlock (2002) calls “intuitive politicians” since we always have to defend ourselves against accountability pressures from others. The main function of reason, therefore, might be primarily argumentative, serving the function of presenting convincing and justifiable arguments for our decisions, actions and beliefs (Mericer & Sperber, 2011). Since we appear to care more about justifiability than about truth, moral reasoning is many times biased and drives people towards decisions that they can justify, even if they are not always optimal (Mercier & Sperber, 2011). The wide use of moral rationalization indicates that there might be a strong sequential relationship between intuition and reasoning, where, as Haidt (2001) claims, intuition comes first and strategic reason second. Further, this type of reasoning should have little impact on moral judgement and behaviour, since it is mainly a post-hoc rationalization that occurs after a judgement or a decision has been made.

The other kind of moral reasoning appears to have a somewhat different relationship with the intuitive system. When we engage in this type of moral reasoning, we do not simply reach a judgement through conscious mental activity, but we reach it in a principled way, in order to be consistent with our other moral commitments (Paxon & Greene, 2010). Tatlock (1999) found that this form of reasoning is especially likely when people are unconstrained (emotionally detached) by past commitments and when we have to explain ourselves to a powerful and knowledgeable audience that have a legitimate right to question our beliefs our decisions. In other words, we seem to need a strong motivation to engage in this type of reasoning.

Parts of the brain associated with principled moral reasoning are the DLPFC and the bilateral inferior parietal lobe (Greene et al, 2004). The anterior DLPFC is associated with abstract reasoning and cognitive control (Greene et al, 2004). Greene (et al, 2004) argues that this part of the brain seems to compete with emotional responses in difficult moral dilemmas, and when it “wins” tend to produce utilitarian judgements. In other words, the relationship between moral reason and moral intuition in the production of a moral judgement might not always be sequential but also competitive. People with a more “rational” and intellectual style, as well as patients with emotional deficits, tend to give more utilitarian answers to moral dilemmas (Bartels, 2008: Moore, Clark & Kane 2008: Koenigs et al, 2007). These results do in some ways support Greene's (et al, 2004) dual process theory in that moral reasoning (utilitarian cost-benefit analysis) is subserved by the DLPFC and tend to prevail when there is no prepotent emotional response. However, Greene's (et al, 2004) dual process theory about competitive systems has been challenged by Moll and de Oliveira-Souza (2007). They claim that in order for a dualistic theory of emotion and cognition (or reason) to work, one would have to prove that selective damages to the DLPFC would lead to more emotional choices while damage to the VMPFC would lead to more utilitarian choices. Although several studies support the claim that VMPFC-damages tend to produce more utilitarian judgements, there is insufficient evidence supporting that DLPFC-damages tend to produce more "emotional" judgements. Disruption of the DLPFC has shown to produce more utilitarian judgement (Tassy et al, 2012), which does not fit with Greene's (et al, 2004) dualistic theory. Further, the lateral part of the DLPFC has also been associated with more self-focused emotions like anger and frustration (Moll & de Oliveira-Souza, 2007).

As an alternative to Greene's (et al, 2004) dual process theory, Moll and de Oliveira-Souza (2007) emphasizes the degree of integration of emotional and cognitive mechanisms as crucial for prosocial moral sentiments. And if one lacks this integration, then there is no strong emotional response in personal dilemmas, hence a utilitarian bias. This theory would explain why VMPFC-patients are more utilitarian when utilitarian judgements are self-serving. However, a clear self-serving motive in making a utilitarian judgement to a hypothetical moral dilemma seems vague. Further research is needed in order to specify the exact role of the DLPFC and how it relates to more "emotional" parts of the brain in the production of a moral judgement.

Two things are important to point out from this discussion. Greene (2007) argues that the kind of cognition that is used for unbiased and principled moral reasoning is capable of emotionally neutral representations. These representations do not trigger particular behavioural responses while "emotional" representations, on the other hand, have behavioural effects. But can one really claim that we are able to produce a consequentialist judgement that accept killing one person in order to save five if we did not have a sentiment that valued human life? If some "cognitive" representations were neutral, then they would not motivate us to do anything. To claim that it is better that one dies instead of five is dependent on that we place value on other people's existence. And value seems to come from emotions. Greene (2007) is aware of this problem but claims that the emotions involved in consequential judgements are different from the emotions that are triggered in personal dilemmas. The former function more like a currency and the latter function more like an alarm (Greene, 2007). What he means by this is that the alarming emotions give simple commands like: " Don't do it!" or "Must do it!" while the more subtle emotions are the basis for the currency of human life that we take into account when we engage in cost-benefit analyses of how to maximize human well-fare (Greene, 2007). Consequentialist judgements are in other words not emotionless, but the process is systematically different because it can be used to dominate decisions and judgement rather than merely influence it.

Let's say that we have an ability to engage in principled moral reasoning that tend to produce utilitarian answers to moral dilemmas. Why do we not use it more often? And how does utilitarian judgement relate to moral behaviour? Principled moral reasoning is by its nature systematic and aggregative, where we have to take into account as much information as we possibly can (Greene, 2007). Everything therefore becomes a complex guessing game where judgements can be changed by additional details (Greene, 2007). Engaging in this kind of reasoning is energy-consuming (Kahneman, 2013) and we have a limited working memory that cannot store and process infinite information (Damasio, 1994). Using this form of reasoning before making a moral judgement or engaging in moral behaviour would therefor require a strong motivation to do so. Further we would probably not get much done besides getting lost in never ending cost-benefit analyses. Therefore, for practical reasons, intuitions will probably always guide and affect our judgements and behaviour in the moral domain, and in other domains.

The second question concerning how utilitarian judgement relates to moral behaviour is extensive and will not be addressed in detail. However, it is interesting to point out that people that tend to give utilitarian answers to difficult moral dilemmas are people with emotional deficits who lack pro-social sentiments (Moll & de Oliveira Souza, 2007). Psychopathy is a personality disorder characterized by a lack of empathy and guilt or remorse, irresponsibility, shallow affect and poor behavioural control (Kiehl, 2008). While psychopaths are capable of articulate morally appropriate responses to moral dilemmas, their actions in real life are frequently inconsistent with their verbal reports (Kiehl, 2008). It appears that good action requires both sincere moral judgement and moral motivation. And when engaging in principled moral reasoning that overrides intuitive responses, we might create a similar dissociation between moral judgement and moral motivation that are seen in psychopaths. An interesting study by Schwitzgebel & Rust (2014) investigated whether moral philosophers specialized in ethics behaved better or at least more consistent with their expressed values than controls. However, on no issue did they show better behaviour than controls, and inconsistencies between expressed values and behaviour were strong when it came to charitable donations. Maybe principled moral reasoning can change our judgement but it might also tend to create dissociation between moral judgement and moral motivation.

This dissociation is apparently not exclusive to psychopaths. In times of environmental threats, poverty and access to weapons of mass destruction, principled moral reasoning may allow us to make judgements concerning what we should do to handle these global issues. However, the challenge will be to back these judgments up with moral motivational mechanisms that seem to have been shaped by evolution to enable successful cooperation within groups but not between groups (Greene, 2013).

References

Adolphs, R., Gosselin, F., Buchanan, T. W., Tranel, D., Schyns, P. & Damasio, A. R. (2005). A mechanism for impaired fear recognition after amygdala damage. *Nature,* 433, 68-72.

Algoe, S.B. & Haidt, J. (2009). Witnessing excellence in action: the 'other-praising' emotions of elevation, gratitude, and admiration. *The Journal of Positive Psychology*, 4 (2), 105-127.

Baron-Cohen, S. (2012). *Zero Degrees of Empathy: A New Theory of Human Cruelty and Kindness.* Great Britain: Penguin Books

Bartels, D. M. (2008). Principled moral sentiment and the flexibility of moral judgment and decision making. *Cognition,* 108 (2), 381–417.

Batson, C. D., Chao, M. C. & Givens, J. M. (2009). Pursuing moral outrage: Anger at tourture. *Journal of Experimental Social Psychology*, 45 (1), 155-160.

Batson, C. D., Kennedy, C. L., Nord, L. A., Stocks, E. L., Fleming, D. A., Marzette, C. M., Lishner, D. A., Hayes, R. E., Kolchinsky, L. M. & Zerger, T. (2011). Anger at unfairness: Is it moral outrage? *European Journal of Social Psychology*, 37, 1272–1285.

Baumeister, R. F., Stillwell, A.M. & Heatherton, T. F. (1994). Guilt: An Interpersonal Approach. *Psychological Bulletin*, 115 (2), 243-267.

Beer, J. S., Heerey, E. A., Keltner, D., Scabini, D. & Knight, R. T. (2003). The regulatory function of self-conscious emotion: Insights from patients with orbitofrontal damage. *Journal of Personality and Social Psychology*, 85(4), 594-604.

Bechara, A., Damasio, H., Damasio, A.R. (2003). Role of the amygdala in decision-making. *Annals of the New York Academy of Science*, 985, 356-369.

Blair, R.J.R. (1995). A cognitive developemental approach to morality: Investingating the psychopath. *Cognition*, 50 (1), 1-29.

Churchland, P. S. (2011). *Braintrust: What Neuroscience Tells Us about Morality*. Princeton, New Jersey: Princeton University Press.

Damasio, A. R. (1994). *Descartes' error: Emotion, Reason and the Human Brain*. New York:

G.P. Putnam.

David, B., & Olatunji, B. O. (2011). The effect of disgust conditioning and disgust sensivity on appraisals of moral transgression. *Personality and Individual Differences*, 50 (7), 1142‑1146.

De Waal, F. B. M. (2008). Putting the altruism back into altruism: the evolution of empathy. *Annual Review of Psychology*, 59, 279–300.

Dougherty, D. D., Shin, L. M., Alphert, N. M., Pitman, R. K., Orr, S. P., Lasko, M., Macklin, M. L., Fischman, A. J. & Rauch, S.L (1999). Anger in healthy men: a PET study using script-driven imagery. *Biological Psychiatry*, 46 (4), 466-472.

Dunbar, R. (1996). *Grooming, gossip, and the evolution of language.* Cambridge, MA: Harvard University Press.

Dunbar, R.I.M. (2004) Gossip in Evolutionary Perspective. *Review of General Psychology*, 8 (2), 100-110.

Ekman, P. (1992) Are there basic emotions? *Psycholosical Review*, 99 (3), 550-553.

Englander, Z. A., Haidt, J. & Morris, J. P. (2012). Neural Basis of Moral Elevation Demonstrated through Inter-Subject Synchronization of Cortical Activity during Free-Viewing**. Retrieved May 13, 2014 from PLOS ONE Journal**: http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0039384#pone-0039384-g004.

Evans, J. S. B. T. (2008). Dual-Processing Accounts of Reasoning, Judgement, and Social Cognition. *Annual Review of psychology*, 1, 255-278.

Frith, C. D. & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, 50, 531–534.

Funk, C. M. & Gazzaniga, M. S. (2009). The Functional Brain Architecture of Human

Morality. *Current Opinion in Neurobiology*, 19 (6), 678‑681.

Galotti, K. M. (1989). Approaches to Studying Formal and Everyday Reasoning. *Psychological Bulletin*, 105(3), 331-351.

Gazzaniga, M. S. (2000). Cerebral specialization and interhemispheric communication. Does the corpus callosum enable the human condition? *Brain*, 127 (7), 1293-1326.

Graybiel, A.M. (2008). Habbits, rituals and the evaluative brain. *Annual Review of Neuroscience*, 31, 359-387.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M. & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral Judgment. *Science*, 293, 2105-2108.

Greene, J. D. (2007). The Secret Joke of Kant's Soul. In W. Sinnot-Armstrong (Eds.), *Moral Psychology, Vol. 3: The Neuroscience of Morality: Emotion, Disease, and Development* (pp. 35-79). Cambridge, MA: MIT Press.

Greene, J. (2013). *Moral Tribes: Emotion, Reason and the Gap Between Us and Them*. New York: The Penguin Press.

Greenwald, A. G., & Banaji, M.R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102 (1), 4-27.

Haidt, J., Koller, S. & Dias, M. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology*, 65, 613–628.

Haidt, J., Rozin, P., McCauley, C. & Imada, S. (1997). Body, psyche and culture: the relationship between disgust and morality. *Psychology and Developing Societies*, 9, 107–13.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgement. *Psychological Review,* 108, 814-834.

Haidt, J. (2003). The moral emotions. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.). *Handbook of Affective Sciences* (pp. 852-870). Oxford: Oxford University Press.

Haidt, J. (2008). Morality. *Perspectives on Psychological Science*, 3(1), 65-72.

Haidt, J. (2012). *The Rightous Mind: Why Good People are Divided by Politics and Religion*. New York: Pantheon.

Hauser, M. (2006). *Moral minds: How nature designed our universal sense of right and wrong*. New York: Harper Collins.

Harris, L. T., & Fiske, S. T. (2007). Social groups that elicit disgust are differentially processed in the mPFC. *Social Cognitive Affective Neuroscience*, 2, 45–51.

Hume, D. (1969-1970). *A Treatise of Human Nature*. London: Penguin.

Hutcherson, C. A. & Gross, J. J. (2011). The Moral Emotions: A Social–Functionalist

Account of Anger, Disgust, and Contempt. *Journal of Personality and Social Psychology*, 100 (4), 719-737.

Kahneman, D. (2011). *Thinking Fast and Slow*. New York: Farrar, Straus and Giroux.

Kant, I. (1785/2002). *Groundwork for the metaphysics of morals*. New Haven, CT: Yale University Press.

Kiehl, K.A. (2008). Whithout Morals: The Cognitive Neuroscience of Criminal Psychopaths. In W. Sinnot-Armstrong (Eds.), *Moral Psychology, Vol. 3: The Neuroscience of Morality: Emotion, Disease, and Development* (pp. 119-149). Cambridge, MA: MIT Press.

Koenigs, M. & Tranel, D. (2007). Irrational economic decision-making after ventromedial prefrontal damage: evidence from the Ultimatum Game. *Journal of Neuroscience*, 27, 951‑956.

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M. & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature*, 446, 908–911.

Kohlberg, L. & Gilligan, C. (1971). The Adolescence as a Moral Philosopher: The discovery of the self in a postconventional world. *Daedalus*, 100 (4), 1051-1086.

Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, 125, 255–275.

Mallon, R. & Nichole, S. (2011). Dual processes and moral rules. *Emotion Review*, 3 (3), 284-285.

McCollough, M. E., Kilpatrick, S. D., Emmons, R. A. & Larsson, D. B. (2001) Is Gratitude a Moral Affect? *Psychological Bulletin*, 127(2), 249-266.

Mercier, H. & Sperber, D. (2011) Why do Humans Reason? Arguments for an Argumentative theory. *Behavioural and Brainscience*, 34, 57–111.

Miller, R. S. (1996*). Embarrassment: Poise and Peril in Everyday Life*. New York: Guilford

Press.

McGeer, V. (2008) Varieties of Moral Agency: Lessons from Autism (and Psychopathy). In W. Sinnot-Armstrong (Eds.), *Moral Psychology, Vol. 3: The Neuroscience of Morality: Emotion, Disease, and Development* (pp. 227-257). Cambridge, MA: MIT Press.

Miller, W. I. (1997). *The anatomy of disgust*. Cambridge, MA: Harvard University Press.

Moll, J., de Oliveira-Souza, R., Eslinger, P. J., Bramati, I. E., Mourao-Miranda, J., Andreiuolo, P. A. & Pessoa, L. (2002). The neural correlates of moral sensitivity: a functional magnetic resonance imaging investigation of basic and moral emotions. *The Journal of Neuroscience,* (22), 2730– 2736.

Moll, J., de Oliveira-Souza, R., Moll, F., Ignácio, F., Bramati, I., Caparelli-Dáquer, E. & Elsinger, P. (2005a). The Moral Affiliations of Disgust: a Functional MRI Study. *Cognitive and Behavioural Neurology*, 18(1), 68-78.

Moll, J., Zahn, R., De Oliveria-Souza, R., Krueger, F. & Grafman, J. (2005b). The Neural Basis of Human Moral Cognition. *Nature Reviews of Neuroscience*, 6, 799-809.

Moll, J. & De Oliveria-Souza, R. (2007). Moral judgement, emotions and the utilitarian brain. *Trends in Cognitive Neuroscience*, 11 (8), 319-321.

Moll, J., De Oliveira-Souza, R. & Zahn, R. (2008). The Neural Basis of Moral Cognition: Sentiments, Concepts and Values. *Annals of the New York Academy of Science,* 1124, 161-180.

Moll, J. & Schulkin, J. (2009). Social attachment and aversion in human moral cognition. *Neuroscience & Biobehavioral Reviews*, 33 (3), 456-465.

Moore, A., Clark, B., & Kane, M. (2008). Who shalt not kill? Individual differences in working memory capacity, executive control, and moral judgment. *Psychological Science,* 19(6), 549–557.

Ovid. (1986). *Metamorphoses*. Transl. Melville, A. D. New York: Oxford University Press Inc.

Paxton, J. M., Greene, J. D. (2010). Moral reasoning: Hints and allegations. *Topics in Cognitive Science*, 2(3), 511-527.

Paxton, J. M., Ungar, L. & Greene, J. D. (2011). Reflection and reasoning in moral judgment. *Cognitive Science*, 36(1), 163-177.

Prinz, J. (2007). *The Emotional Construction of Morals*. New York: Oxford University Press Inc.

Robbins, T. W. & Arnsten, A. F. T. (2009). The neuropsychopharmacology of fronto-executive function: monoaminergic modulation. *Annual review of neuroscience,* 32(1), 267-287.

Rozin, P & Fallon, A.E. (1987). A perspective on disgust. *Psychological Review*, 94 (1), 23-41.

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(4), 574–586.

Russell, P. S. & Roger, G. S. (2013). Bodily Moral Disgust: What It Is, How It Is Different From Anger and Why it is an Unreasoned Emotion. *Psychological Bulletin*, 2 (139), 328-351.

Sambataro, F., Dimalta, S., Di Giorgio, A., Taurisano, P., Blasi, G., Scarabino, T., Giannatempo, G., Nardini, M. & Bertolino, A. (2006). Preferential responses in amygdala and insula during presentation of facial contempt and disgust. *European Journal of Neuroscience,* 24, 2355–2362.

Schachter S. & Singer, J.E. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review,* 69, 379–99.

Scherer, K. R. (1997). The role of culture in emotion-antecendent appraisal. *Journal of Personality and Social psychology*, 73, 902-922.

Schnall, S., Haidt, J. Clore, G. L & Jordan, A. H. (2008). Disgust as embodied moral Judgement. *Personality and Social Psychology Bulletin*, 34(8), 1096-1109.

Sherman, G.D. & Haid, J. (2011). Cuteness and Disgust: The Humanizing and Dehumanizing Effects of Emotion. *Emotion Review*, 3 (3), 245-251.

Schwitzgebel, E. & Rust, J. (2014). The Moral Behaviour of Ethics Professors. Relationships among self-reported behavior, expressed normative attitude, and directly observed behavior. *Philisophical Psychology*, 27 (3), 293-327.

Singer, P. (1972). Famine, Affluence and Morality. *Philosophy and Public Affairs*, 1(1), 229-243.

Singer, P. (1985/2011). *The Expanding Circle: Ethics, Evolution and Moral Progress.* Princetown and Oxford: Princetown Univerisy Press.

Singer, T., Kiebel, S. J., Winston, J. S, Dolan, J. R. & Frith, C. D. (2004). Brain Responses to the Acquired Moral Status of Faces. *Neuron*, 41, 653-662.

Singer, T. (2006). The neuronal basis and ontogeny of empathy and mind reading: Review of Literature and implications for future research. *Neuroscience & Behavioural Reviews*, 30, (6), 855-863.

Singer, T., Seymour, B., Doherty, J. P., Stephan, K. E., Dolan, R. J. & Frith, C. D. (2006). Empathic neural responses are modulated by the percieved fairness of others. *Nature,* 439 (7075), 466-469.

Singer, T. & Lamm, C. (2009). The social neuroscience of empathy. *Annals of the New York Academy of Science*, 1156, 81-96.

Stark, R., Zimmermann, M., Kagerer, S., Schienle, A., Walter, B., Weygandt, M. & Vaitl, W. (2007). Hemodynamic brain correlates of disgust and fear ratings. *NeuroImage*, 37, (2), 663-673.

Takahashi, H., Yahata, N., Koedad, M., Matsuda, T., Asai, K. & Okubo, Y. (2004). Brain activation associated with evaluative processes of guilt and embarrassment : an fMRI study. *NeuroImage,* 23 (3), 967-974.

Tangney, J.R. (1991). Moral affect: the good, the bad and the ugly. *Journal of Personality and Social Psychology,* 61 (4), 598-607.

Tangney, J. P. & Miller, R. S. (1996). Are shame, guilt and embarrassment distinct emotions? *Journal of Personal and Social Psychology*, 70 (6), 1256-1269.

Tangney, J. P., Stuewig, J. & Mashek, D. J. (2007). Moral emotions and moral behavior. *Annual Review of Psychology*, 58, 345-372.

Tangney, J. P., Stuewig, J., Mashek, D. & Hastings, M. (2011). Assessing Jail Inmates' Proneness To Shame and Guilt. Feeling Bad About the Behavior or the Self? *Criminal Justice and Behavior,* 38 (7), 710-734.

Tassy, S., Oullier, O., Duclos, Y., Coulon, O., Mancini, J., Deruelle, C., Attarian, S., Felician, O. & Wicker, B. (2012). Disrupting the Right Prefronta Cortex Alters Moral Judgement. *Social Cognitive and Affective Neuroscience*, 7, (3), 282-288.

Tetlock, P. E., Skitka, L., & Boettger, R. (1989). Social and cognitive strategies for coping with accountability: Conformity, complexity, and bolstering. *Journal of Personality and Social Psychology*, 57, 632–640.

Tetlock, P.E. (2002). Social Functionalist Frameworks for Judgement and Choice: Intuitive Politicians, Theologians, and Prosecutors. *Psychological Review*, 109 (3), 451-471.

Wheatley, T. & Haidt, J. (2005). Hypnotic disgust makes moral judgements more severe. *Psychological Science,* 16 (10), 780-784.

Woodward, J. & Allman, J. (2007). Moral intuition: Its neural substrates and normative significance. *Journal of Physiology-Paris*, 101 (4-6), 179-202.