The Neural Basis of Morality

Maarten A.S. Boksem & David De Cremer

The interest in ethical, and especially unethical behaviour is rapidly increasing as witnessed by the many articles published in magazines and newspapers on fraud, corruption and bad CEO behaviour. To understand how people come to violate ethical and moral norms, it is important to arrive at a better understanding of how and why people come to behave in an ethical manner in the first place. In this chapter, we will show that moral behaviour is deeply rooted in our biological system. Throughout our evolution as human beings, the social group that we belong to has played a role of central importance: belonging to a well-functioning social group is essential for the survival of the individual. It is therefore not surprising that we have developed specialized biological and neural systems that allow us to function successfully within these social groups. These neural systems are proposed to form the basis of what we now call moral behaviour: taking into account the interests and integrity of individuals other than ourselves and the strong disapproval of behaviour that would harm others or would compromise our group.

In this chapter, we will explore how neuroscience methods have been applied to investigating moral behaviour. The overall aim of the present chapter is to present a short overview of recent neuroscientific research examining the processes and neural substrates involved in ethical decision-making and linking it to concepts of prosocial emotions and the biological basis of basic attachment between individuals. This may provide us with a better understanding of the origins and function of moral behaviour.

This chapter was published in Psychological Perspectives on Ethical Behavior and Decision Making, pages 153–166.

1. Moral evaluations and the Prefrontal Cortex

In recent years, functional magnetic resonance imaging (fMRI) has been employed to examine the neural substrate involved in the processing of emotional valence and moral content in social judgments. Using a strong magnetic field and high frequency radio waves, fMRI makes it possible to image both the anatomy, as well as the activation of specific brain structures. Basically, an fMRI scanner can detect local differences in blood oxygenation, which is a direct correlate of neural activity (active brain cells (neurons) require more oxygen than inactive neurons).

One of the first studies to convincingly show a neural network underlying moral evaluations was that of Moll and colleagues (2002). While in the scanner, subjects were requested to judge statements from three main conditions: emotionally unpleasant statements with moral connotations (‘He shot the victim to death’), emotionally unpleasant statements without moral connotations (‘He licked the dirty toilet’), and emotionally neutral statements (‘He never uses a seat belt’). Subjects were asked to classify these statements as either right or wrong. Comparing brain activation elicited by unpleasant statements with or without moral connotation, the authors found increased activation in Brodmans Areas (BA) 10 and 11 (the anterior part of the medial orbitofrontal cortex (OFC)). This area of the brain has previously been implicated in estimating the outcome value of behavioural actions, suggesting that this brain area may also be critically involved in integrating moral knowledge with the emotions that determine the value of these actions (Moll et al., 2002).

Another early study took a slightly different approach, by scanning the brains of subjects solving moral dilemmas (Greene et al., 2001). One such dilemma is the trolley dilemma: A runaway trolley is speeding down the track towards five people who will be killed if it continues on its present course. The only way to save these people is for you to throw a switch that will diverge the trolley onto an alternate set of tracks where it will kill one person instead of five. Ought you turn the trolley in order to save five people at the expense of one? The footbridge dilemma poses a similar problem: as before, a trolley is bound to kill five people. You are standing next to a rather large stranger on a footbridge that spans the tracks, in between the oncoming trolley and the five people. In this scenario, the only way to save the five people is to push this stranger off the bridge, onto the tracks below. He will die if you do this, but his body will stop the trolley from reaching the others. Ought you save the five others by pushing this stranger to his death? While most people answer yes to the first dilemma, they say no to the second.

The authors argue that the crucial difference between these dilemmas lies in the fact that the footbridge dilemma tends to engage people's emotions in a way that the trolley dilemma does not. The thought of physically pushing someone to his death, they propose, is more emotionally and morally salient than the thought of operating a switch that will cause similar consequences. Supporting their hypothesis, Greene and colleagues (2001) found that the more personal dilemmas (like the footbridge dilemma) was associated with greater activation of brain areas similar to these reported by Moll and colleagues (BA 9 and 10; medial prefrontal cortex (mPFC); a brain area that is known to be involved in emotional processes), compared to the more impersonal dilemmas such as the trolley dilemma.

These, and other studies, converge on the fact that the medial areas of the prefrontal cortex (PFC) are primarily involved in moral evaluations. This is not to say that other areas are not involved. Indeed, also lateral areas of the PFC have been shown to be activated by moral tasks (Berthoz et al., 2002; Finger et al., 2006), as well as the amygdala (Moll et al., 2002; Greene et al., 2004; Berthoz et al., 2006), the cingulate cortex (Greene et al., 2001; Moll et al, 2002), superior temporal sulcus (Heekeren et al., 2003, 2005), and other brain areas, making moral evaluations a distributed, whole brain affair. However, the mPFC has been focused on the most in research on morality and we will therefore mainly limit our discussion to this area of the brain.

Recently, Koenigs and colleagues (2007) investigated the behaviour of patients with lesions to exactly this area of the brain, the ventro-medial PFC (vmPFC) on dilemmas such as those used by Greene and colleagues (2001). They found that, whereas normal subjects are likely to say that it is allright to throw a switch, killing one instead of five, but do not endorse pushing a bulky stranger onto the tracks to stop the train, vmPFC lesioned subjects judged both cases to be acceptable. Like Greene and colleagues (2001), these authors argued that it is the emotional salience of the actions involved in the more 'personal' moral dilemmas (actually pushing another human being of a bridge) that is integrated into the decision-making process by the vmPFC that leads normal subjects to reject such actions. Patients with vmPFC lesions, by consequence fail to involve these emotional considerations in their decisions, making them more 'utilitarian' (favouring the larger aggregate welfare over the welfare of fewer individuals). Although this is a tempting interpretation (emotional blunting is commonly observed in patients with frontal brain damage) it seems to be at odds with a second study by Koenigs and Tranel (2007).

In that study, patients with vmPFC damage played the 'ultimatum game'. In a typical instantiation of this game, two players are given one opportunity to split a sum of money. One player (the proposer) offers a portion of the money to the second player (the responder). The responder can either accept the offer (in which case both players split the money as proposed) or reject the offer (in which case both players get nothing). If the responder would act strictly utilitarian, he or she should accept any offer, no matter how low, because rejecting the offer would mean getting nothing at all. However, relatively small offers (20% of the total) have a 50% chance of being rejected (Bolton and Zwick, 1995). The "irrational" rejection of unfair Ultimatum Game offers has been correlated with feelings of anger (Pillutla and Murnighan, 1996), increased skin conductance responses (van 't Wout et al., 2006), and activation of the insula (Sanfey et al., 2003), a brain area associated with negative emotional states (Phillips et al., 1997). Therefore, lacking such a strong emotional reaction to unfair offers, one would expect patients with vmPFC damage to no longer irrationally reject these unfair offers. Surprisingly, what Koenigs and Tranel (2007) found was that vmPFC patients rejected unfair offers even more often that control subjects did. These patients allowed the negative emotions provoked by receiving such an unfair offer to overrule their utilitarian reason more than normal subjects would do. This makes the emotional blunting in these patients seem like an unlikely explanation for the findings in the moral dilemma study by Koenigs et al. (2007).

A more plausible interpretation of these findings was provided by Moll and deOlivera-Souza (2007). They suggested that the vmPFC may be involved in the experience of prosocial moral sentiments, such as guilt, compassion and empathy, and the integration of these emotions with cognitive mechanisms. These prosocial sentiments are proposed to emerge when emotional states mediated by the limbic system, are integrated with mechanisms mediated by more anterior sectors of the PFC, such as prospective evaluation of action outcomes. This integration of emotions and cognition would allow us to assess the consequences of our own actions upon the wellbeing of others (Moll et al., 2005; Moll and deOlivera-Souza, 2007).

While vmPFC is suggested to be involved in the experience of prosocial moral sentiments, more lateral areas of the PFC, such as dorsolateral PFC and lateral OFC are more important for self-centred and other-aversive emotional experience (e.g. anger, frustration or moral disgust) (Moll et al., 2002; 2005; 2006). Selective damage to the vmPFC, along with more lateral areas being spared, may explain the otherwise puzzling discrepancy in the findings in the two studies by Koenigs and colleagues. The vmPFC-damaged patients playing the ultimatum game let emotions such as anger and frustration guide their actions towards the non-utilitarian decision to reject unfair offers. On the other hand, vmPFC patients were more utilitarian when solving moral dilemmas, because the damage to this part of their prefrontal cortex reduced their prosocial sentiments, steering them towards decisions that favour aggregate total welfare over the wellbeing of individuals.

Prosocial, or moral emotions such as guilt, shame, compassion, promote the helping of others, cooperation and conforming to social norms, probably by invoking feelings of social attachment (Eisenberg, 2000). They function to regulate social behaviours, often in the long-term interests of a social group rather than the short-term interests of the individual person (Haidt, 2001). The importance of actually feeling these emotions is painfully clear from observations of people who apparently do not have these feelings. Psychopathy, or sociopathy, is a type of antisocial personality that physicians in the 19th century labelled 'moral insanity' (Berrios, 1999). It is a developmental disorder (Lynam et al. 2007) that involves emotional dysfunction, characterized by reduced guilt, empathy and attachment to significant others. As Blair (2007) puts it: 'These people can steal from their friends, dismember live animals, and even murder their parents to collect insurance benefits without showing any trace of remorse or, when caught, of shame. Psychopaths know the rules of social behaviour and they understand the harmful consequences of their actions for others. They simply do not care about those consequences'.

Neuroimaging studies on psychopathy have found that individuals with psychopathy show reduced activation of both the amygdala and the rostral anterior cingulate cortex/vmPFC, compared to normal subjects (Kiehl et al., 2001; Birbaumer et al., 2005). Although damage to specific brain areas is not associated with psychopathy, damage to ventral, medial and polar aspects of the frontal cortex results in severe personality changes and behaviour very similar to that observed in psychopaths. This condition has been called acquired sociopathy (Anderson, et al., 1999) and is associated with injury to specific brain areas, particularly the orbitomedial and polar frontal cortex, the medial frontal lobe, and several basal forebrain structures (Adolphs *et al.,* 1998; Blair and Cipolotti, 2000). Like primary psychopaths, such acquired sociopaths often remain perfectly able to tell right from wrong and to articulate sound statements on morality and social appropriateness, that stand in sharp contrast to their actual behaviour.

2. Prosocial behaviour, basic attachment and their neural correlates

The fact that psychopaths are unable to function in our society, shows us just how important prosocial moral sentiments are for a successful life in a social context. Indeed, helping others has been proposed to be evolutionary favourable: individuals that cooperate when the helping of others provides long term benefits over competing with others have better chances of survival and procreation. There has been a long debate on how helping others, often at one's own expense (altruism), would be evolutionary adaptive. A big step in modelling the evolution of helping behaviour is the extension of reciprocal altruism ("I'll help you if you help me") by "indirect reciprocity" (Nowak and Sigmund, 2005) in which prosocial behaviour pays off by improving one's reputation, which elicits cooperation from others at a later point in time. Reputation is a powerful force for the development and consolidation of social communities. When players know each other's reputations in repeated-play behavioural economics games, cooperation rates improve substantially (Fehr and Henrich, 2003). Evolutionary models show that indirect reciprocity can solve the problem of selfish behaviour known as free-riding (which doomed simpler models of altruism) in moderately large groups (Panchanathan and Boyd, 2004), when people have access to information about reputations (Haidt, 2007). Being socially successful by maintaining a good reputation is highly predictive of both chances of survival of the individual and for successful passing on of the genes that lead to this behaviour. By this process, it is very much in the interest of the individual to help others and work for the common interest of the group in order to achieve and maintain a good reputation.

However, as often happens in evolution, the impulse to help close others to maintain good standing in the social group, became separated from the consequences that caused the evolution of this behaviour in the first place. By this process, this impulse also occurred when the chance of reciprocity was slim, so that also strangers could benefit. This is why we usually also treat people we do not really know in a just and ethical manner. It is important to realize that, certainly in evolutionary terms, it is not so long ago that meeting strangers was an extremely rare event. All social interactions were with a very limited number of people that were of the same family, clan or village. That we now live in what has been called a global village is in evolutionary terms a very recent novelty. Our brains evolved under circumstances in which social groups were much smaller and the prosocial emotions generated in our brains still guide us toward behaviour that benefits others that we perceive to be close to us. For example, a huge number of people donated money to help victims of the tsunami in Thailand and Indonesia, whom they had never met and never will meet and who will never be able to return the favour. Moreover, we feel indignation when we hear of innocent people being imprisoned and tortured in some country far away from our own, when it is absolutely clear that those people are completely unrelated to us.

While moral behaviour towards other people in our global society can be viewed as a generalization of prosocial behaviour towards close group members, at a very basic level, the inclination to help next of kin and close others may in itself be a generalization from the basic attachment between mother and infant. In this relationship, it is of course of primary importance for the survival of the infant that the mother not only pursues her own immediate interests, but is also sensitive to the needs of her child. In non-human animals, the neuropeptide oxytocin has a central role in stimulating this maternal care (Pedersen, 1997). In addition, oxytocin receptors are distributed in various brain regions associated with behaviour such as pair bonding, sexual behaviour and the ability to form normal social attachments (Insel and Young, 2001). Given these findings, recent studies have begun to investigate the role of oxytocin in the promotion of prosocial behaviours in humans.

Kosfeld and colleagues (2005) analysed the effect of exogenously administered oxytocin (using a nasal spray) on individuals' decisions in a trust game with real monetary stakes. In this trust game, two subjects interacting anonymously play either the role of an investor or a trustee. First, the investor has the option of choosing a costly trusting action by giving money to the trustee. If the investor transfers money, the total amount available for distribution between the two players is doubled by the experimenter but, initially, the trustee reaps the whole increase. The trustee is then informed about the investor's transfer and can honour the investor's trust by sharing the monetary increase generated by the investor's transfer. Thus, if the investor gives money to the trustee and the latter shares the proceeds of the transfer, both players end up with a higher monetary payoff. However, the trustee also has the option of violating the investor's trust. As sharing the proceeds is costly for the trustee, a selfish trustee will never honour the investor's trust because the investor and the trustee interact only once during the experiment. Administration of oxytocin by Kosfeld and colleagues, increased trust considerably: subjects that were administered oxytocin on average transferred 17% more money to the trustee, with 45% of the treated subjects transferring all of their money, while only 21% of non-treated subjects transferred the entire amount.

In another study, subjects were administered either oxytocin or placebo and played the ultimatum game (Zak et al., 2007). The results showed that oxytocin-administered proposers offered a significantly larger (21%) amount of money to the responders. The authors propose that oxytocin selectively affected the understanding of how the another would experience receiving a stingy offer and were motivated to reduce these expected negative emotions in others.

Although it does not become clear from these studies exactly how oxytocin affects neural processing in terms of enhancing activity in certain brain areas, there is some evidence from animal studies that the mPFC may play a role. For example, microinjection of an OT receptor agonist into the mPFC enhances pair bond formation in female prairie voles (a small rodent living in North America; Young et al., 2001). In addition, the mPFC has been found to integrate information from brain structures involved in empathy and social behaviour and that are rich in OT receptors. The reader will recall that it also was the mPFC that was shown to be activated in studies on human morality (Greene et al., 2002; Moll et al., 2005; Koenigs et al., 2007).

All of this thus suggests that our sense of morality originates from a 'hard wired' neural system that may be based on an evolutionary old system of basic attachment. For the survival of a newborn infant, it is of great importance that the mother feels connected to and responsible for the infant; its interest should be of equal or even greater importance to the mother than her own interests. If this were not the case, the infant would not survive and indeed its entire species would be doomed for extinction. For social animals (including humans), these feelings of attachment are not limited to mother and child, but extend also to next of kin, or even nonrelated group members. As already mentioned, in a social setting it is of primary importance for the individual to maintain good relations with group members: all member are dependent on each other, and in the end chances of survival are greater when the individual not just acts out of immediate self-interest, but instead cooperates with others. Indeed, pure selfish behaviour may lead to social exclusion, and once excluded from the group, chances of survival are often limited. This is why moral behaviour is so fundamental to our being and is associated with such strong emotions. Moral behaviour is prosocial behaviour and prevents exclusion from our social group. Violating moral norms (thus compromising our good reputation) or observing others violate moral norms (thus compromising the integrity of our social group) evokes strong aversive emotions such as shame, indignation or anger. It is these emotions that push our behaviour away from pure selfishness towards behaviour that promotes the wellbeing of others.

3. The role of reasoning: Theory of mind

Although morality or ethics has traditionally been posited to depend on a 'higher', typically human capacity to reason in order to come to the 'right' conclusion regarding how one should act in a given situation, it appears that reasoning may play only a relatively minor role in moral decision making. Indeed, Haidt (2001) argues in his social intuitionist model of morality, that 'moral judgment is much like aesthetic judgment: we see an action or hear a story and we have an instant feeling of approval or disapproval. These feelings are best thought of as affect-laden intuitions, as they appear suddenly and effortlessly in consciousness, with an affective valence (good or bad), but without any feeling of having gone through steps of searching, weighing evidence, or inferring a conclusion. People certainly do engage in moral reasoning, but, as suggested by studies of informal reasoning, these processes are typically one-sided efforts in support of pre-ordained conclusions.

While our reasoning abilities appear to play only a minor role in morality, most people would agree that our moral behaviour is something more than prosocial behaviour in 'lower' mammals or even primates. Although this difference may not be nearly as great as has been accepted until recent years, humans do excel at a form of reasoning that is highly social in nature. Our evolution in social groups has shaped our brain in such a way so that it is particularly good at thinking about what others may be thinking about. This capacity for social-cognitive processing has been proposed to enable us to go from empathy (which has also been observed in other primates and even mammals) to sympathy, which has so far has only been observed in humans

To be able to take the interests of others into account (i.e. display prosocial behaviour), it is fundamental to be able to understand what those interests are, and to be able to realize that these interests may differ from one's own. This capacity has been termed mentalizing or theory of mind (ToM), and involves the process of inferring the feelings and thoughts of others from their observed behaviour (Blakemore and Decety, 2001; Frith and Frith, 1999). The ability to accurately predict what others are thinking is essential for the initiation of social interactions: it enables us to predict how others may react to our actions and what the subjective and emotional consequences of our actions will be for these others.

ToM emerges at around four years of age in young children, as is evidenced by their performance on the so-called 'Sally Ann' task. In this task, infants are shown a scenario in which Sally has a pram, and Ann has a box of toys. In the scenario, Sally puts one of her toys in the pram and then leaves the scene. When Sally has left, Ann takes the toy out of the pram and puts it in her box. After a while, Sally returns for her toy. Where will she look for the toy? Normal children of about four years of age and older correctly answer that she will look in the pram, where she believes the toy to be, although the child knows it is actually in the box. Younger children or children with disorders like autism answer that she will look in the box (where they know the toy actually is, not realizing that Sally's mind, being different from their own, does not contain this information).

Studying ToM in adults is a bit more complex but is generally also done using scenarios or cartoons. Contrasting scenarios in which ToM is necessary to understand what is going on to scenarios for which this is not required results in increased activation in several brain areas, but particularly in the vmPFC (Frith and Frith, 1999; Berthoz et al., 2002). The tendency to attribute intentions and feelings to others is so strong and automatic that we even do it with objects that are clearly non-social. A classic example involves subjects watching a screen on which basic shapes (triangles, squares) moving about. Subjects readily ascribe intentions to these objects like: 'The big red square is trying to persuade the little square to move', or 'The blue triangle is trying to sneak up on the red triangle' (e.g. Heider and Simmel, 1944). Interestingly, watching these scenes activates the same areas that have been shown to be involved in mentalizing or ToM (Castelli et al., 2000).

In recent years, the discovery of so-called 'mirror neurons' has had a large impact on the theory of ToM (see Rizzolatti et al., 2001). The existence of these cells were discovered more or less by accident (or so the story goes). Researchers were studying the neural activity in macaque monkeys involved in grasping and picking up food items. After they were done, they forgot to turn off the recording equipment. At a certain moment, someone walks in with an ice-cream cone. To the surprise of all present, the monkeys 'grasping neurons' began to fire spontaneously without any observed movement. Apparently, these cells simulated the movement of grasping for the ice-cream, without actually doing it.

Of course, instead of offering the monkey some of the ice-cream, this prompted the researchers to submit this monkey to various additional tests, showing that these cells not only fire when the monkey performs a certain motor act, but also when it views someone else performing the action. In particular, it was shown that these cells respond to the *intention* of the observed acts. This mechanism would allow for understanding the intentions of the behaviours of others, by simulating these actions internally. Although probably not representing the complete neural substrate of ToM, coding the intentions behind movements observed in others may be an essential part of what makes full blown ToM possible.

4. Conclusion

Together, our ability to understand the intentions of others and our strong prosocial sentiments that have evolved from basic attachment to important others in our social group, allow for morality to be a driving force in our social setting, guiding behaviour away from pure self-interest and towards interests that we share with significant others. However, it is important to realize that selfishness and prosocial behaviour are not as diametrically opposed as they may seem. As already mentioned, displaying prosocial behaviour is often in the best interest of the individual, increasing chances of a successful life. Therefore, even prosocial behaviour can be seen as being ultimately selfish. Indeed, as becomes clear from reading the newspapers or watching the news on TV, people are often remarkably unconstrained in violating moral norms when individual profits can be made and they feel they can get away with it (i.e. when it will not diminish their social standing). While moral behaviour is aimed at preventing social exclusion, if chances of exclusion are slim (when moral violations will remain undetected), certain people may be less inclined to behave morally when there is a clear self-interest motive. A similar process may occur in situations when the entire group with which individuals identify themselves displays unethical behaviour, as may happen within a bad ethical climate in companies. If certain ethical violations become the norm for a group (of CEOs, for example), individual group members may not feel a risk of exclusion and self-interest motives may become dominant. An extreme example may be wartime situations in which people commit atrocities against other human beings. In these circumstances, in-group vs. out-group sentiments are very strong and amoral behaviour against out-group members (i.e. the enemy) may not be perceived as such because it will not lead to exclusion from the in-group (see also Optow et al., 2005). Indeed, it may even increase one's standing in the in-group.

In summary, we propose that moral behaviour is an evolutionary adaptation to living in a highly social environment wherein taking into account the interests of significant others is of central importance to maintaining good standing in the group, in turn preventing social exclusion. We have suggested that the required prosocial sentiments have evolved from basic attachment mechanisms between mother and infant and have been transferred to close group members and finally also to individuals that are not able or likely to reciprocate prosocial acts. We have shown that the neuropeptide oxytocin may play a central role in prosocial and moral behaviour and that certain brain areas, particularly the vmPFC are primarily involved in moral evaluations and decision making. Through our capacity for ToM, we have the uniquely human ability to display true sympathy for our fellow humans. Together, these evolutionary old neural systems drive ethical behaviour which is still such an important issue in our modern-day society.

5. References

Adolphs, R., Tranel, D., & Damasio, A. R. (1998). The human amygdala in social judgment. *Nature, 393*(6684), 470-474.

Anderson, S. W., Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1999). Impairment of social and moral behavior related to early damage in human prefrontal cortex. *Nature Neuroscience, 2*(11), 1032-1037.

Berrios, G. E. (1999). J. C. Prichard and the concept of 'moral insanity'. *History of Psychiatry, 10*(37), 111-116.

Berthoz, S., Armony, J. L., Blair, R. J. R., & Dolan, R. J. (2002). An fMRI study of intentional and unintentional (embarrassing) violations of social norms. *Brain, 125*, 1696-1708.

Berthoz, S., Grezes, J., Armony, J. L., Passingham, R. E., & Dolan, R. J. (2006). Affective response to one's own moral violations. *Neuroimage, 31*(2), 945-950.

Birbaumer, N., Viet, R., Lotze, M., Erb, M., Hermann, C., Grodd, W., et al. (2005). Deficient fear conditioning in psychopathy - A functional magnetic resonance imaging study. *Archives of General Psychiatry, 62*(7), 799-805.

Blair, R. J. R. (2007). The amygdala and ventromedial prefrontal cortex in morality and psychopathy. *Trends in Cognitive Sciences, 11*(9), 387-392.

Blair, R. J. R., & Cipolotti, L. (2000). Impaired social response reversal - A case of 'acquired sociopathy'. *Brain, 123*, 1122-1141.

Blakemore, S. J., & Decety, J. (2001). From the perception of action to the understanding of intention. *Nature Reviews Neuroscience, 2*(8), 561-567.

Bolton, G. E., & Zwick, R. (1995). Anonymity versus punishment in ultimatum bargaining. *Games and Economic Behavior, 10*(1), 95-121.

Castelli, F., Happe, F., Frith, U., & Frith, C. (2000). Movement and mind: A functional imaging study of perception and interpretation of complex intentional movement patterns. *Neuroimage, 12*(3), 314-325.

Eisenberg, N. (2000). Emotion, regulation, and moral development. *Annual Review of Psychology, 51*, 665-697.

Fehr, E., & Henrich, J., (2003). In P. Hammerstein (Ed), *Genetic and Cultural Evolution of Cooperation*, Cambridge, MA: MIT Press.

Finger, E. C., Marsh, A. A., Kamel, N., Mitchell, D. G. V., & Blair, J. R. (2006). Caught in the act: The impact of audience on the neural response to morally and socially inappropriate behavior. *Neuroimage, 33*(1), 414-421.

Frith, C. D., & Frith, U. (1999). Cognitive psychology - Interacting minds - A biological basis. *Science, 286*(5445), 1692-1695.

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron, 44*(2), 389-400.

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*(5537), 2105-2108.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*(4), 814-834.

Haidt, J. (2007). The new synthesis in moral psychology. *Science, 316*(5827), 998-1002.

Heekeren, H. R., Wartenburger, I., Schmidt, H., Prehn, K., Schwintowski, H. P., & Villringer, A. (2005). Influence of bodily harm on neural correlates of semantic and moral decision-making. *Neuroimage, 24*(3), 887-897.

Heekeren, H. R., Wartenburger, I., Schmidt, H., Schwintowski, H. P., & Villringer, A. (2003). An fMRI study of simple ethical decision-making. *Neuroreport, 14*(9), 1215-1219.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *American Journal of Psychology, 57*, 243–259.

Insel, T. R., & Young, L. J. (2001). The neurobiology of attachment. *Nature Reviews Neuroscience, 2*(2), 129-136.

Kiehl, K. A., Smith, A. M., Hare, R. D., Mendrek, A., Forster, B. B., Brink, J., et al. (2001). Limbic abnormalities in affective processing by criminal psychopaths as revealed by functional magnetic resonance imaging. *Biological Psychiatry, 50*(9), 677-684.

Koenigs, M., & Tranel, D. (2007). Irrational economic decision-making after ventromedial prefrontal damage: Evidence from the ultimatum game. *Journal of Neuroscience, 27*(4), 951-956.

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., et al. (2007). Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature, 446*(7138), 908-911.

Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U., & Fehr, E. (2005). Oxytocin increases trust in humans. *Nature, 435*(7042), 673-676.

Lynam, D. R., Caspi, A., Moffitt, T. E., Loeber, R., & Stouthamer-Loeber, M. (2007). Longitudinal evidence that psychopathy scores in early adolescence predict adult psychopathy. *Journal of Abnormal Psychology, 116*(1), 155-165.

Moll, J., & de Oliveira-Souza, R. (2007). Moral judgments, emotions and the utilitarian brain. *Trends in Cognitive Sciences, 11*(8), 319-321.

Moll, J., de Oliveira-Souza, R., Bramati, I. E., & Grafman, J. (2002). Functional networks in emotional moral and nonmoral social judgments. *Neuroimage, 16*(3), 696-703.

Moll, J., de Oliveira-Souza, R., Moll, F. T., Ignacio, F. A., Bramati, I. E., Caparelli-Daquer, E. M., et al. (2005). The moral affiliations of disgust - A functional MRI study. *Cognitive and Behavioral Neurology, 18*(1), 68-78.

Moll, J., Krueger, F., Zahn, R., Pardini, M., de Oliveira-Souzat, R., & Grafman, J. (2006). Human fronto-mesolimbic networks guide decisions about charitable donation. *Proceedings of the National Academy of Sciences of the United States of America, 103*(42), 15623-15628.

Nowak, M. A., & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature, 437*(7063), 1291-1298.

Opotow, S., Gerson, J., & Woodside, S. (2005). From moral exclusion to moral inclusion: Theory for teaching peace. *Theory into Practice, 44*(4), 303-318.

Panchanathan, K., & Boyd, R. (2004). Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature, 432*(7016), 499-502.

Pedersen, C. A. (1997). Oxytocin control of maternal behavior - Regulation by sex steroids and offspring stimuli. In *Integrative Neurobiology of Affiliation* (Vol. 807, pp. 126-145).

Phillips, M. L., Young, A. W., Senior, C., Brammer, M., Andrew, C., Calder, A. J., et al. (1997). A specific neural substrate for perceiving facial expressions of disgust. *Nature, 389*(6650), 495-498.

Pillutla, M. M., & Murnighan, J. K. (1996). Unfairness, anger, and spite: Emotional rejections of ultimatum offers. *Organizational Behavior and Human Decision Processes, 68*(3), 208-224.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience, 2*(9), 661-670.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science, 300*(5626), 1755-1758.

van't Wout, M., Kahn, R. S., Sanfey, A. G., & Aleman, A. (2006). Affective state and decision-making in the Ultimatum Game. *Experimental Brain Research, 169*(4), 564-568.

Young, L. J., Lim, M. M., Gingrich, B., & Insel, T. R. (2001). Cellular mechanisms of social attachment. *Hormones and Behavior, 40*(2), 133-138.

Zak, P.J., Stanton, A.A., Ahmadi, S. (2007). Oxytocin Increases Generosity in Humans. *PLoS ONE, 2*(11), e1128.