From actions to empathy and morality – A neural perspective

Istvan Molnar-Szakacs a, b, c, *

a Tennenbaum Center for the Biology of Creativity, University of California, Los Angeles, CA 90095, USA
b Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, CA 90095, USA
c Semel Institute for Neuroscience and Human Behavior, David Geffen School of Medicine, University of California, Los Angeles, CA 90095, USA

A R T I C L E   I N F O

Article history:
Received 15 July 2009
Received in revised form 25 February 2010
Accepted 27 February 2010
Available online 16 September 2010

Keywords:
Mirror neuron system
Limbic system
Emotion
Empathy
Morality
Neuroimaging

A B S T R A C T

Our culturally varied, complex social world, governed by unwritten moral codes that encourage affiliative helping behavior, may be subserved by the unique properties of a neural system for understanding the intentions and actions of others. The firing pattern of neurons within this system appears to ‘mirror’ an action performed and seen, coding a functional correspondence between a motor action and sensory perception of that action. Indirect evidence acquired through various neuroimaging techniques supports the presence of such a neural system, termed the mirror neuron system (MNS) in the human brain. In this paper I discuss evidence suggesting that the human MNS – by linking intention and outcome, perception and action, observer and actor – forms part of the neural system for empathic concern, the capacity to understand and feel another’s emotional state. By helping to establish a ‘likeness’ between interacting agents, the human MNS may support the active desire to understand others, to feel what they are feeling and to help alleviate another’s suffering. By providing a biological substrate for such fundamental affiliative behaviors, the MNS may provide a neural scaffold for the evolution of our sophisticated sociality and the morality that governs it.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

A neural system implementing a resonance mechanism exists in the primate brain, allowing for shared representations to be established between interacting agents. At the neural level, this resonance mechanism in essence ‘mirrors’ one agent to another, this function providing the name to the mirror neuron system (MNS). Mirror neurons were first discovered and described in the ventral premotor area F5 of the macaque monkey brain (Gallese et al., 1996). While it is beyond the scope of the current review to detail the properties of the MNS in the monkey (see Fogassi, this issue), it suffices to say that the MNS of an agent encodes the intentions, actions and emotions of another as if of the self, without a distinguishing boundary between self and other. For instance, you might jerk your leg and feel your calves flex if you saw your friend get kicked in the shins during a soccer match – as the same neural activation would be triggered in your brain by observing his accident, as if you were to have experienced it yourself! This triggered ‘likeness’, your feeling that it could have been you who experienced the injury, allows you to understand your friend’s pain, and this empathic feeling may even motivate you to run to the field and help him if he fell to the ground. Thus, our capacity for empathy – the ability to share another’s emotions and feelings – is based in a simple ‘mirroring’ mechanism implemented by the human MNS that allows us to use the same neural resources to represent states of the self and others in an overlapping way.

* Corresponding author at: Tennenbaum Center for the Biology of Creativity, Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles, 760 Westwood Plaza, Room C8-846, Los Angeles, CA 90095, USA. Tel.: +1 310 206 4666; fax: +1 310 206 8525.
E-mail address: imolnar@ucla.edu.

0167-2681/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jebo.2010.02.019
The aim of the current paper is to review and summarize recent research on the human MNS, asserting that this system may provide the neural scaffold for empathy and, as a result, may have served as a biological substrate for the evolution of our sophisticated sociality and the morality that governs it. Section 1 serves as the general introduction, Section 1.1 discusses neuroimaging evidence in support of the human MNS; Section 1.2 highlights the role of the MNS in understanding the intentions of others; Section 2 discusses a form of empathy termed ‘emotional empathy’, that is the feeling of other’s feelings, putatively through a resonance mechanism implemented by the human MNS; Section 2.1 discusses the relationship between ‘emotional empathy’ and ‘cognitive empathy’ – a form of mentalizing that relies on higher-level cognitive processes; Section 3 discusses developmental evidence on social evaluation and empathy for pain; Section 4 aims to link action perception to morality, using emotional empathy and social evaluation as stepping stones, and Section 5 concludes.

1.1. The human mirror neuron system

Soon after the discovery of mirror neurons in the monkey brain using single-unit electrode recordings (di Pellegrino et al., 1992; Gallese et al., 1996; Rizzolatti et al., 1996a), non-invasive neuroimaging (Grafton et al., 1996; Rizzolatti et al., 1996b) and transcranial magnetic stimulation (TMS) studies (Fadiga et al., 1995) demonstrated a network with similar functional properties in the human brain. Fadiga et al. (1995) provided an elegant demonstration of the matching neural representation in humans, for actions observed and those performed by the self. In their study, magnetic stimulation was delivered to the scalp above the motor cortex – the source of motor commands for action: (1) while participants were observing an experimenter perform various hand actions in front of them, or (2) during control conditions including arm movement observation, object observation and dimming detection, while motor evoked potentials (MEPs) were recorded from hand muscles. MEPs serve as a quantifiable index of activity within the motor cortex. Results showed that during hand action observation, but not in the other conditions, there was an increase in the amplitude of the MEPs in the same hand muscles that are used when the observed action is actually performed by the observer (Fadiga et al., 1995). This increase in MEPs resulting from observing a hand action reflects a facilitation, or priming of the motor cortex – an increased potential for action – due to the neural ‘mirroring’ of the hand action that was observed. Subsequent work then confirmed and extended these findings (Borroni and Baldiessera, 2008; Gangitano et al., 2001; Montagna et al., 2005; Strafella and Paus, 2000), leading researchers to conclude that in humans there is a neural system, resembling the one described in the monkey, matching observed actions and executed actions.

Imitation is a ubiquitous human behavior and learning mechanism for language and culture, composed of both an observation and an execution phase, making it an ideal paradigm for the study of neural activity associated with observation–execution matching in the human brain. Functional Magnetic Resonance Imaging (fMRI) was used to localize brain regions matching observed action to executed action using an imitation paradigm. Based on their predictions on neural firing rates in the monkey (Gallese et al., 1996), Iacoboni et al. (1999) hypothesized that areas of the human brain that show mirror properties would show an increase in brain activity in the same area during action observation and during action execution, and action execution would lead to approximately double the increase in activity, as action execution contains an overt motor response. In turn, imitation, which contains both an observation and execution component would lead to the greatest increase in neural activity. This study found two cortical areas that showed this predicted pattern of activity, the posterior inferior frontal gyrus (IFG) and the rostral sector of the inferior parietal lobule (IPL) (Iacoboni et al., 1999).

In addition to the early neuroimaging experiments described above, a wealth of additional studies have shown that mirror resonance mechanisms are involved in action perception and performance, meaning that overlapping neural systems are recruited during the observation and execution of actions (Aziz-Zadeh et al., 2006; Binkofski et al., 1999; Buccino et al., 2004; Fadiga et al., 1995; Grafton et al., 1996; Grezes et al., 2003; Iacoboni et al., 2005, 1999; Johnson-Frey et al., 2003; Koski et al., 2002; Molnar-Szakacs et al., 2005; Nishitani and Hari, 2000; Oberman et al., 2005; Rizzolatti et al., 1996b). Taken together, results of these studies imply that perception and action are inseparable, both in our daily behavior, as well as within the neural networks supporting them in our brain. Thus, the human brain contains a neural system for representing actions both observed and executed, similar to that described in the monkey, formed by the inferior premotor cortex and the inferior parietal lobule (Rizzolatti and Craighero, 2004).

1.2. Intention understanding and the mirror neuron system

The fundamental starting point of any action or behavior is the intention that gives rise to it – and the human MNS has been proposed as a neural instantiation of such motor intentions (Iacoboni et al., 2005). A recent neuroimaging study using fMRI has shown that the human MNS also encodes the perceived intentions that give rise to overt motor actions. In this experiment, participants were simply instructed to observe hand actions embedded within different contexts, as well as hand actions without a context (Iacoboni et al., 2005). These conditions provided information about the action alone (no context) or both the action and intention (deduced from the contextual information). Results showed that observation of actions embedded within a context led to more activity in the MNS than actions without a context, and different contexts (implying different intentions) led to differential levels of activation of the MNS (Iacoboni et al., 2005). These findings imply that the human MNS is sensitive to the context within which actions are embedded, and codes actions that imply different
intentions differently, providing support for the hypothesis that human mirror areas mediate the understanding of the intentions behind other's actions. These results are in line with the data recently reported on intention coding by parietal mirror neurons in the monkey as well (Fogassi et al., 2005).

Another interesting manipulation in this experiment was that half of the participants were simply told to watch the actions, while the other half were given explicit instructions to watch and to consciously try and deduce the intention behind the observed actions. When data from these two groups was compared, observation of actions with instructions to simply watch versus instructions to infer intention yielded no differences in activity within the MNS. This finding showed that the representation of intentions and actions is automatic and pre-cognitive – the MNS is activated without conscious effort (Iacoboni et al., 2005). This implies that “mirror neurons are a kind of ‘neural wi-fi’ that monitors what is happening in the other people. This system tracks their emotions, what movements they’re making, what they intend and it activates, in our brains, precisely the same brain areas as are active in the other person. This puts us on the same wavelength and it does it automatically, instantaneously and unconsciously” (Goleman, 2006).

Thus, the MNS may be able to establish a representational equivalence between perception and action at the neural level, may provide the basis for representational equivalences between self and other on a behavioral level, enabling more profound types of social relatedness, such as empathy (Gallese, 2003b).

To further explore intention understanding in a social context, we used single-pulse TMS delivered to the primary motor cortex of participants, to probe brain activity during the observation of either an ethnic ingroup member or outgroup member performing culturally familiar or foreign gestures (Molnar-Szakacs et al., 2007). It has been theorized that “the more similar and socially close two individuals are, the easier the subject’s identification with the object” (de Waal, 2008). In other words, communication at the behavioral level and resonance at the neural level, would be facilitated for targets similar to the agent, and less effective for targets dissimilar from the agent. Thus, we predicted that a Caucasian American participant would be able to ‘mirror’ the actions of an ethnic ingroup member (Caucasian American) relatively more readily than those of an ethnic outgroup member (Hispanic Nicaraguan). As predicted, we found that Caucasian American participants showed stronger motor resonance (higher MEPs), while observing gestures performed by a Caucasian than a Hispanic actor. However, activity within the MNS to specific gestures was modulated by the communicative intent of those gestures. Meaningful Nicaraguan gestures elicited stronger motor resonance in the MNS of the American participants than meaningless Nicaraguan gestures, even if technically, both types of gestures were foreign and meaningless to our participants. Thus, beyond the influence of cultural familiarity, the MNS was able to extract the intentions behind communicative versus meaningless gestures (Molnar-Szakacs et al., 2007). While these results were the first to provide evidence that the human MNS may be sensitive to ethnic group membership, it seems that it is the intention behind the gestures that may be the most subtle, yet important aspect of social interaction coded by the human MNS, a result consistent with our previous neuroimaging findings (Iacoboni et al., 2005).

2. Emotional empathy and the mirror neuron system

In addition to representing the motor intentions and actions of others, the human MNS may also contribute to the representation of the emotional states of others, in order to understand and predict their behavior (Adolphs, 2009). A neural mechanism like the MNS, that uses our own representations of intentions and actions in order to understand the same in others is the ideal neural substrate for empathy, defined as the inter-personal sharing of experiences, emotions and feelings. In fact, empathy may be considered a special case of a more general state-matching reaction (resonance) that is fast, automatic and pre-cognitive. When Lipps (1903) introduced the concept of empathy (Einfühlung), he theorized that ‘inner imitation’ of the actions of others plays a critical role in establishing a relationship between self and other.

In a fMRI study of this hypothesis, participants were asked to either observe or imitate emotional facial expressions (Carr et al., 2003). If ‘inner imitation’ of the observed facial expression of others is truly a component of the empathic response, then the mere observation of emotional facial expressions should activate the human MNS. Thus, observation and imitation of emotional facial expressions should yield substantially similar patterns of activated brain areas. The results of the study confirmed the hypothesis, showing a similar network of activated areas – including the premotor face area, the posterior IFG, the superior temporal sulcus, the insula, and the amygdala – for both imitation and observation of emotional facial expressions (Carr et al., 2003). The results showed that, in fact, observation of another person in a particular emotional state automatically activates a representation of that state in the observer, as was originally proposed by Lipps (1903). These data also provide support for the notion that empathy is a special form of resonance, whereby the emotional state of one individual is reflected in another, recruiting not only the human MNS, but also regions of the limbic system, important for the processing of emotion.

The limbic system – a term often used to describe regions that constitute the ‘emotional brain’, includes both cortical regions like the orbito-frontal cortex (OFC), the anterior cingulate cortex (ACC), the ventromedial prefrontal cortex (vmPFC) and the anterior insula, and subcortical structures including the amygdala, the nucleus accumbens (NA) and the hypothalamus – is critical for emotional behavior, and the representation of emotions, including pain (Pessoa, 2008). Anatomical data suggest that a sector of the insular lobe, the dysgranular field, is connected with the limbic system as well as with posterior parietal, inferior frontal, and superior temporal cortices constituting the human MNS (Augustine, 1996). This connectivity
pattern makes the insula an ideal functional link between the motor representation functions of the MNS and the emotion representation functions of the limbic system, and an integral part of the neural network for empathy involved in representing and understanding affective states (Craig, 2002, 2003; Iacoboni, 2009; Pessoa, 2008). In fact, the anterior insula is critically involved in interoceptive processing, and the consolidation of bodily state information about oneself with social and contextual information from the environment to provide a neural substrate for the conscious experience of emotions (Craig, 2008).

Preston and de Waal formalized a theory of emotional-motor resonance in the Perception-Action Model (PAM), which holds that perception of a behavior performed by another automatically activates one's own representations for the behavior, and output from this shared representation automatically proceeds to motor areas of the brain where responses are prepared and executed. This basic emotional contagion, or resonance, supported by interactions between the MNS and the limbic system, is fast, automatic and pre-cognitive, and is thought to support our ability to empathize emotionally (I feel what you feel) (Preston and de Waal, 2002). Neuroimaging studies have provided evidence in support of the model, showing common neural signatures while experiencing disgust (Wicker et al., 2003), touch (Keysers et al., 2004) or pain (Jackson et al., 2006; Singer et al., 2004) in oneself, and when perceiving the same feelings in others. More specifically, studies on empathy for pain have found that the activations observed in mirror regions and the limbic system to experiencing pain are also seen when observing pain in someone else (Singer et al., 2004, 2006).

In parallel with resonance processes occurring at the neural level, humans have a propensity to automatically and unconsciously align their behavior to cues in their environment or to individuals around them – a type of behavioral resonance. For example, when watching your favorite soccer team, you might find your leg muscles tensing at the moment of a shot toward the goal. Furthermore, empathic individuals tend to exhibit non-conscious mimicry of the postures, mannerisms, and facial expressions of others to a greater extent than non-empathic individuals, a behavioral inclination aptly named the chameleon effect (Chartrand and Bargh, 1999). In a clever experimental demonstration of this propensity, a group of participants were exposed to words typically associated with the elderly, such as ‘Florida’, ‘bingo’, ‘gray’, and an experimenter timed their walk as they left the lab. These primed participants walked significantly slower compared to participants who were not exposed to elderly stereotype words, in an example of behavioral resonance, or ‘alignment’ to the elderly stereotype (Bargh et al., 1996). In an automatic and unconscious way, the primed subjects imitated the slowness of the elderly. Several other studies (Dijksterhuis, 2005; Dijksterhuis et al., 2000; Dijksterhuis and van Knippenberg, 1998) have since demonstrated a similar tendency for people to naturally align their behavior with others or environmental cues.

Empathic emotional attunement appears to rely on the direct link between perception and action instantiated by the human MNS. As perceiving an action activates the same representations as performance of the same action, this overlap might allow humans to ‘embody’ the behavior of others and to infer their internal states, including the intentions and emotions driving it (Wilson and Knoblich, 2005). The automatic and pre-conscious re-presentation of the intentions, actions and emotions of another do not require that we explicitly think about them – our brain has a built-in mechanism for knowing them, automatically and unconsciously, as we know our own intentions, actions and emotions. In other words, humans embody the knowledge that other individuals have intentions that are manifest through bodies that think, feel and move in similar ways to their own.

2.1. Cognitive empathy

Current evolutionary evidence suggests that in addition to the phylogenetically early emotional empathy system described above, there is also a more advanced cognitive perspective-taking system mediating empathic responses in humans (de Waal, 2008). Cognitive empathy requires that we actively think about, or reflect on others' actions and emotional states, which requires effortful, higher-level cognitions, including perspective-taking or mentalizing (de Waal, 2008). This process of understanding another person’s perspective appears to depend upon higher cognitive functions such as cognitive flexibility (Decety and Jackson, 2004). Singer (2006) has proposed that cognitive empathy, mentalizing, and ‘Theory of Mind’ (ToM), refer to our ability to understand mental states such as intentions, goals and beliefs, while emotional empathy refers to our ability to share the feelings of others. Neuroimaging studies have implicated distinct neural networks subserving cognitive versus emotional empathic responses (Hooker et al., 2008; Shamay-Tsoory et al., 2004; Singer, 2006; Vollm et al., 2006) – a ventromedial prefrontal cortex (vmPFC) – temporal lobe network supporting cognitive empathy (mentalizing, ToM, perspective-taking), while emotional empathy is thought to rely mostly on resonance processes implemented by the MNS–limbic system network (Gallese, 2005; Iacoboni, 2009; Iacoboni and Dapretto, 2006; Preston and de Waal, 2002).

To test a true dissociation between these two proposed empathic systems, groups of subjects with lesions in either the vmPFC (part of the cognitive empathy system) or inferior frontal gyrus (IFG) (part of the emotional empathy system) and two control groups were assessed with the Interpersonal Reactivity Index (IRI; Davis, 1983), a measure of empathy that incorporates both cognitive and emotional dimensions. A cognitive item to be evaluated with reference to one’s own behavior would be: ‘I sometimes try to understand my friends better by imagining how things look from their perspective’; whereas an example of an emotional dimension item would be: ‘I often have tender, concerned feelings for people less fortunate than me’ (Davis, 1983). The findings revealed a remarkable behavioral and anatomic double dissociation, whereby patients with vmPFC lesions were impaired in cognitive empathy and patients with IFG lesions were impaired in emotional empathy. Precise anatomical mapping of the brain damage revealed the IFG to be critical for emotional empathy while the vmPFC was found necessary for cognitive empathy (Shamay-Tsoory et al., 2009).
Thus, there are at least two ways to come to understand others' intentions and behavior – a bottom-up, unconscious and automatic resonance mechanism, and a top-down, effortful, cognitive mentalizing mechanism – supported by distinct neural systems that often interact. In fact, it has been suggested that one may consider the outputs of the resonance mechanism as the potential inputs to the cognitive mentalizing system (Adolphs, 2009). We may first generate motor representations of another person's actions or emotions (via resonance) and then use this representation in more flexible ways to predict future behaviors taking into consideration context, culture and beliefs (via mentalizing) (Keyser and Gazzola, 2007).

In an elegant demonstration of the interplay of emotional and cognitive empathy, Singer et al. (2006), had participants in a neuroimaging study observe others receiving painful stimulation, and found empathy-related brain activation within the fronto-insular and anterior cingulate cortices – a response subserved by the emotional empathy system. A clever manipulation in the experiment, set up some of the ‘observed others’ to be perceived as fair or unfair individuals. Interestingly, the neural response to observing others receive painful stimulation was modulated by the perceived fairness of the person receiving pain, such that empathy-related responses were significantly reduced in males when observing an unfair person receiving pain, relative to a fair person. This modulation of the empathic response is likely due to a higher level, cognitive evaluation of the other’s fairness, and demonstrates how interactions between the emotional and cognitive empathy systems can give rise to complex social evaluations. Furthermore, the observed decrease in empathic response was accompanied by increased activation in reward-related areas, correlated with an expressed desire for revenge. The authors concluded that their results offer a neural foundation for theories of social preferences (Fehr and Schmidt, 1999; Rabin, 1993), which suggest that people value the gains of others positively if they are perceived to behave fairly, but value such gains negatively if they behave unfairly. This pattern of preferences implies that people not only like cooperating with fair opponents but also like punishing unfair opponents (Singer et al., 2006). These data show, that the automatic representation of the intentions behind other’s actions, as well as the higher level social evaluation that follows, are key components of the empathic response, showing the inextricable link between automatic resonance-based emotional empathy and higher level, cognitive empathy. Thus, the seemingly most complex aspects of our socio-emotional life, our relationships and interactions are likely constructed from fundamental elements based in neural mechanisms of action perception and production (Gallese, 2007; Iacoboni, 2009).

3. Development of social evaluation and empathy

Is the ability to feel the joy and pain of another by simply observing the expression of those emotions a biological adaptation or a learned ability? Hamlin et al. (2007) tried to get closer to the answer by having infants 6–10 months old observe a character, the ‘climber’, at rest at the bottom of a hill. During a habituation phase, infants saw the climber repeatedly attempt to climb the hill, and on the third attempt, was either aided up by a ‘helper’ who pushed it from behind, or was pushed down by a ‘hinderer’. The researchers then asked how infants as observers, evaluate helpers and hinderers. Will witnessing one individual’s actions towards an unknown third party affect the infants’ attitude towards that individual? To test this, infants were encouraged to choose between the helper or hinderer. Implicit in this choice, is the assumption that infants can automatically and unconsciously deduce the intention of the climber to get to the top of the hill, and the intention of the ‘helper’ to aid him, and the ‘hinderer’ to prevent him. Using looking times as a choice measure to indicate preference, the 10-month-old infants robustly chose the helper, indicating that they held distinct impressions of the two characters on the basis of their actions towards the climber (Hamlin et al., 2007). The infants demonstrated a preference for those who act cooperatively to facilitate the goals of others, and do not prefer those who impede another’s goals – a developmental parallel to the neuroimaging results of Singer et al. (2006) with ‘fair’ and ‘unfair’ others. Hamlin et al. (2007) concluded that the ability to judge differentially those who perform positive and negative social acts may form an essential basis for more abstract concepts of right and wrong. The preverbal infants in this study showed not only that they automatically inferred the intention of the ‘climber’ to ascend the slope, but also the social intentions of the ‘helper’ and the ‘hinderer’ towards the climber. The presence of such social evaluation this early in infancy suggests that assessing individuals by the nature of their interactions with others is not only central to processing the social world, but also an automatic, effortless process. Both evolutionarily and developmentally, this capacity may serve as the foundation for moral thought and action, and its early developmental emergence supports the view that social evaluation is indeed a biological adaptation, rather than an acquired skill.

Asking a fundamentally similar question about empathy in a neuroimaging study, children aged 7–12 years observed others experience accidental pain, as well as one person intentionally inflicting pain onto another (Decety et al., 2008). For both conditions, increases in neural activity were observed in regions of the limbic system involved in the processing of first-hand experience of pain, including the ACC, the somatosensory cortex, the SMA, periaqueductal gray and the insula. Interestingly, when watching a person intentionally inflict pain on another, regions that are engaged in mentalizing and moral behavior – the temporoparietal junction, the paracingulate, medial orbitofrontal cortices (mOFC) and amygdala – were additionally recruited, and increased their connectivity with the fronto-parietal attention network (Decety et al., 2008). Connectivity analyses of neuroimaging data are used to estimate the functional coupling of different brain regions that work together, such that increased connectivity indicates a strong functional association, whereas low connectivity can reflect weak interaction or functional dysregulation between regions. ‘Connectivity’ in this case refers to functional, rather than anatomical connectivity among interacting brain regions, determined using psychophysiological interaction (PPI) analyses of co-varying activity (Friston et al., 1997). Taken together, these results provide both behavioral and biological evidence
in developing populations for the empathic response and accompanying automatic social evaluation that occurs when observing an individual intentionally harm another, as well as evidence for an overlap in the neural substrates of observing and experiencing pain, supporting the notion that these foundational elements of moral behavior were selected for through evolution.

4. From actions to empathy to morality

We have built a case for the human MNS as a neural system for coding actions, observed and executed, as well as the coding of motor intentions behind those actions. Through interactions between the MNS and the limbic system, this automatic resonance mechanism allows one to 'feel what another feels', while higher level cognitive empathy adds layers of information about other's perspective and mental state, leading to complex social evaluations such as 'helping/harming' or 'fairness'. Both ontogenetically and phylogenetically, the capacity for empathy scaffolds social evaluation, which may serve as the foundation for moral thought and action.

Empathy has been considered central to the human moral affective system for at least three reasons (Eisenberg et al., 2004, 2006). First, as we have discussed, empathic reactions to others’ distress often elicit feelings of concern for the distressed other (Feshbach, 1975). This has been noted by many philosophers and even made the central focus of Adam Smith’s theory of morality (Smith, 1759). At the neural level, brain regions representing one’s own feelings and emotions largely overlap with those representing the feelings and emotions of others (Keysers et al., 2004; Singer et al., 2004; Wicker et al., 2003). Second, empathic concern towards others and the corresponding neural response are modulated by perceived fairness of the other (Hamlin et al., 2007; Singer et al., 2006) and often prompts behavior aimed at helping the distressed other (Batson, 1991; Eisenberg and Miller, 1987). Finally, feelings of empathy are apt to inhibit aggression and other behaviors that are harmful to others (Decety et al., 2009; Feshbach and Feshbach, 1969). Thus, a key aspect of moral thought is the evaluation of one’s own thoughts and actions in relation to others, the social world and cultural norms. Indeed, moral behavior refers to what individuals should do based on principles (i.e. moral values) shared with other members of their social environment.

Haidt has advanced an elaborate account of moral emotions, also referred to as social or sociomoral (for reviews see Greene and Haidt, 2002; Haidt, 2001, 2007). Whereas basic emotions (like happiness, sadness, anger, and so on) are characterized by immediate personal relevance, and spring most directly from perceptions and to a certain extent from ideas, imagination or memories, moral emotions are imbued with social relevance to the interest or welfare either of society as a whole or persons in our social environment (Haidt, 2007; Tangney et al., 2007). Therefore, moral emotions are evoked by circumstances that extend beyond the immediate sphere of the self. Guilt, gratitude and compassion are examples of prosocial moral emotions, while contempt, indignation and xenophobia (a fear or dislike of that which is different from oneself) are antisocial moral emotions.

In one of the first experiments to investigate brain responses to moral emotions, participants passively viewed pictures depicting moral violations (e.g., physical assaults, poor children abandoned in the streets, war scenes; Haidt, 2003), non-moral unpleasant scenes (e.g., body lesions, dangerous animals, body products; Rozin et al., 1999), and unemotional scenes including people and landscapes (Moll et al., 2002b). While both moral and non-moral conditions activated limbic regions, in agreement with previous studies of unpleasant emotion processing, regions important for mentalizing were preferentially recruited by scenes of moral violation. A functional connectivity analysis showed strong functional coupling between these regions, supporting the hypothesis of a specific functional network underlying moral appraisals (Moll et al., 2002b). Thus, spontaneous moral appraisals appear to be subserved by a neural network that in part overlaps the limbic system, including the amygdala and insula, as well as regions important for mentalizing (cognitive empathy) and social evaluation including the mOFC, vmPFC and STS (Moll et al., 2003).

Several other functional neuroimaging studies in healthy individuals have since investigated moral judgments (Greene et al., 2001; Harenski and Hamann, 2006; Heekeren et al., 2003; Moll et al., 2002a, 2007, 2005, 2001) and moral reasoning (Borg et al., 2006; Greene et al., 2008, 2004), and have detected consistent activations of the OFC and vmPFC, although activated regions encompass a more extended network of neural regions including the insula, limbic regions and the temporal poles (Moll et al., 2003; Raine and Yang, 2006; Young and Koenigs, 2007).

As empathy for others is considered an essential component of moral behavior, a lack of empathy has been proposed as a core deficit in youth with aggressive conduct disorder (CD). CD is characterized by a longstanding pattern of violations of rules and laws, including aggression, lying, theft and inflicting pain on others. This propensity for aggressive behavior in individuals with CD has been hypothesized to reflect a blunted empathic response to the suffering of others (Blair, 2005; Lovett and Sheffield, 2007). An alternative hypothesis suggests that empathic mimicry in conjunction with poor emotion regulation might produce negative affect that leads to aggression (Campbell, 1990; Gill and Calkins, 2003). To test these alternative hypotheses, Decety et al. (2009) imaged the brains of adolescents with aggressive CD and matched controls with no CD symptoms, while watching people experiencing pain or not experiencing pain. As before, these situations involved either an individual whose pain was caused by accident or an individual whose pain was inflicted intentionally by another person. After having their brain activity imaged, participants rated how painful the situations were (Decety et al., 2009).

In both groups the perception of others in pain was associated with activation of limbic regions for pain (Decety et al., 2008, 2009). Most interesting however, is that these regions, including the anterior insula, were activated to a greater extent in youth with CD than in typical youth, and they also showed significantly greater amygdala, and temporal pole activation, but decreased activations in regions important for social evaluation including the tempo-parietal junction and mOFC.
Importantly, the OFC plays a key role in emotion regulation and modulating limbic activity (Davidson et al., 2000; Izquierdo et al., 2005), and damage to this area is associated with impulsive and aggressive behavior and limited awareness of the moral implications of one’s actions (Anderson et al., 1999; Grafman et al., 1996). Thus, intact fronto-limbic connectivity is essential for regulating behavior, and youth with CD exhibited less functional fronto-limbic coupling when watching pain inflicted by another than did control youth (Decety et al., 2009). This functional dysregulation of the fronto-limbic pathways in youth with CD may prevent them from fully comprehending the moral consequences of their destructive acts. Thus, rather than a lack of empathy, these results support the theory of a heightened empathic response combined with insufficient emotion regulation leading to the aggressive behavior seen in youth with CD (Decety et al., 2009).

Consistent with these functional neuroimaging results, studies of clinical populations, including patients with vmPFC damage, have revealed an association between impairments in emotional processing and impairments in moral judgment and behavior. Ciaramelli et al. (2007) tested 7 patients with lesions in the vmPFC (part of the cognitive empathy system) and 12 healthy individuals in personal moral dilemmas (i.e. having to decide whether or not to push a stranger off of a footbridge in front of an oncoming trolley in order to save five people on the track), impersonal moral dilemmas (i.e. having to decide whether or not to hit a switch that will turn the trolley to an alternate set of tracks, where it will kill one person instead of five), and non-moral dilemmas (Greene et al., 2001). Compared to the normal controls, patients with damage to their cognitive empathy system were more willing to judge personal moral violations as acceptable behaviors, and they did so more quickly. In contrast, their performance in impersonal and non-moral dilemmas was comparable to that of controls. These results led the authors to conclude that the vmPFC is necessary to oppose personal moral violations, possibly by mediating self-focused, emotional reactions that may exert strong influence on moral choice and behavior (Ciaramelli et al., 2007).

But how are these responses to moral violations related to functioning of the MNS? It has been suggested that psychopathy and antisocial personality disorder (APD) may be characterized by aggression and a lack of guilt and empathy (APA, 2000; Hare, 1985). Fecteau et al. (2008) investigated whether psychopathic personality traits in a healthy adult sample were related to motor components of empathy subserved by the human MNS. The participants viewed short videos of a needle penetrating a human hand, like those that have previously been shown to activate the sensorimotor MNS for pain (Avenanti et al., 2005), and TMS-induced MEPs were recorded as a measure of motor cortex excitability. Individual psychopathic personality traits were assessed using the Psychopathic Personality Inventory (PPI) (Lilienfeld and Andrews, 1996) and correlated with neural data. Observation of the painful stimuli was associated with a significant reduction in the amplitude of the TMS-induced MEPs. Interestingly, the level of corticospinal excitability modulation was positively correlated with individual scores on the coldheartedness subscale of the PPI, such that individuals who scored highest on the coldheartedness measure showed the greatest MEP reduction. These results from a healthy adult sample provide evidence for a functional link between motor components of empathy and psychopathy (Fecteau et al., 2008). In line with the clinical data of Decety et al. (2009), these results suggest that individuals with psychopathological tendencies may show intact empathic responsiveness, and motor components of this response are directly encoded by the MNS.

The evidence gathered here provides support for a shared neural system for experiencing and understanding bodily and emotional states in others. Basic sensori-motor resonance mechanisms subserved by the MNS play a critical role in scaffolding empathy and social evaluation that rely on understanding other’s intentions and actions and the ability to share in others’ distress. Moral decision-making is supported by a healthy functional coupling between the MNS, limbic system and the mentalizing system. Indeed, moral behavior refers to what individuals should do based on principles shared with other members of their socio-cultural milieu. These data suggest that morality is not based in abstract thought, but emerges out of embodied inter-personal sharing, at both neural and behavioral levels.

5. Conclusion

With the goal of tracing a route from simple perception-action mechanisms implemented by the human MNS to empathy and the social nature of moral reasoning and behavior, I have tried to build a case for understanding our seemingly complex socio-moral behavior in terms of its fundamental building blocks – our physical, bodily interactions with the world and our peers. The human MNS provides a neural substrate through which we are able not only to understand the intentions and actions of others, but build social networks through inter-action. The MNS, working in concert with the limbic system, imbibes our thoughts, perceptions and actions with emotional color, allowing us to feel the joys and pains of others as our own. Empathic concern, our capacity to understand and feel another’s emotions, often gives rise to the active desire to alleviate another’s suffering and become a more compassionate human. The mentalizing system regulates the expression of such emotions and actions within our socio-cultural milieu to give rise to complex socio-moral behaviors including helping and harming, fairness and inequity, reward and punishment.

Acknowledgements

I am grateful to the organizers of the Symposium on Emotions, Natural Selection and Rationality, Elias L. Khalil and Jonathan Wight, the participants and attendees of the Symposium, as well as the University of Richmond for its generous support towards making it all possible. I am thankful to Leonardo Fogassi and Katherine Karlsgodt for insightful comments on earlier versions of this manuscript.
References


