I FEEL YOUR PAIN:
MIRROR NEURONS, EMPATHY, AND MORAL MOTIVATION

Paul Thagard
University of Waterloo
pthagard@uwaterloo.ca
Draft 4, August 11, 2006

ABSTRACT: Mirror neurons are brain systems found in monkeys and humans that respond similarly to actions and to the perception of actions of others. This paper explores the implications of mirror neurons for several important philosophical problems, including knowledge of other minds, the nature of empathy, and moral motivation. It argues that mirror neurons provide a direct route to other minds, empathy, and moral motivation that complements the more familiar route based on conscious, verbal inference.

INTRODUCTION

The discovery of mirror neurons has been hailed as one of the major recent breakthroughs in neuroscience, with possible implications for the explanation of many important cognitive functions, including action understanding, imitation, language, and empathy. Mirror neurons were first identified in the 1990s by Giacomo Rizzolatti and his colleagues at the University of Parma (Rizzolatti and Craighero, 2004). They found that monkey prefrontal cortex contains a particular class of neurons that discharge both when the monkey does a particular action and when it observes another individual doing a particular action. Similar classes of neurons have been found in humans, capable of mirroring not only physical actions but also pain and disgust.

This paper is an exploration of the epistemological and ethical significance of mirror neuron mechanisms. After briefly reviewing the recent neuroscience literature on

mirror neurons, I will discuss its implications for the traditional philosophical problem of other minds. Mirror neurons seem to offer a more direct route to the understanding of other people than provided by the usual kinds of inference – analogical and explanatory – that have been considered to provide solutions to the problem of other minds. Similarly, mirror neurons shed light on the nature of empathy, with implications for the moral psychology of caring. Finally, I argue that mirror neurons help with the problem of moral motivation, which concerns the relation between moral judgments and people's willingness to act on them.

MIRROR NEURONS

When a monkey grasps an object, there are neurons in area F5 of its premotor cortex that fire. Much more surprising is the serendipitous discovery by Rizzolatti and his colleagues that the same region contains neurons that fire both when the monkey grasps an object and when it observes another monkey or human grasping an object (Di Pellegrino et al., 1992; Gallese et al., 1996; Rizzolatti et al., 1996; Nelissen et al., 2005). There are mirror neurons in F5 for grasping both with hands and with mouths, and another area, the superior temporal sulcus, contains mirror neurons for walking, turning the head, bending the torso, and moving the arms. The observations represented by mirror neurons are visual-motor, integrating the visual and motor experiences of monkeys. Rizzolatti and Craighero (2004) argue that the mirror neuron system is the basis for both action understanding and imitation. Not only does a monkey's mirror neuron system give it a direct understanding of what another monkey is doing when it moves, but also facilitates imitating those motions that might be useful for its own goals such as finding food. Mirror neurons can also work with auditory-motor representations:

Kohler et al. (2002) found neurons in monkey premotor cortex that discharge when the animal performs a specific action and when it hears the related sound.

The evidence for mirror neurons in monkeys comes from direct recording of single neurons, but evidence for analogous systems in humans is largely indirect, from brain scanning. Rizzolatti and Craighero (2004) cite many studies that show that the observation of actions done by others activates in humans a complex network formed by visual and motor areas. According to Rizzolatti (2005), evidence that a mirror system exists in humans comes from electroencephalography, magnetoencephalography, transcranial magnetic stimulation, and brain imaging studies. Hence observing the physical actions of others prepares people not only to understand what they are doing but also to imitate them. More controversially, Rizzolatti and Arbib (1999) conjecture that mirror-neuron systems provided the neurophysiological basis from which language developed as an extension of gestural communication. Iacoboni et al. (2005) argue that premotor mirror neurons are also involved in understanding others' intentions.

In humans, mirror neurons may be relevant for how people understand emotions as well as actions. Gallese, Keysers, and Rizzolatti (2004) claim that a mirror-neuron system involving visceral-motor centers enables people to understand each other's emotions, just as one involving visual-motor centers enables people to understand each other's actions. Wicker et al. (2003) used fMRI brain scans to compare how people react to disgusting smells with how they react to video clips of people reacting to disgusting smells. They found that the brain's anterior insula, which is known to collect information from various visceral centers, is activated both during the emotion of disgust evoked by unpleasant odorants and during the observation of facial expressions of

disgust. Additional overlap was found in the anterior cingulate cortex. Hence it appears that these two cortical areas, the insula and the anterior cingulate, enable people to directly appreciate other peoples' emotions of disgust.

Similarly, Botvinick et al. (2005) used neuroimaging to find that perception of facial expressions of pain engage cortical areas also engaged by the first-hand experience of pain, including the anterior cingulate and the insula, the same areas that had been found to mirror disgust. Singer et al. (2004, 2006) also found that insular and anterior cingulate cortex were activated both by receiving pain and by observing a loved one receiving pain. Further support for the mirroring of pain is found in the studies of Avenanti et al. (2005) who used transcranial magnetic stimulation to detect evidence for the presence of empathic appreciation of the sensory qualities of the pain of others. I shall discuss later the significance of mirror neurons and empathy and mirror neurons for moral psychology. First, however, I want to explore the relevance of mirror neurons to the traditional philosophical problem of other minds.

OTHER MINDS

There are at least two problems of other minds. The first, philosophically skeptical one, concerns how you can be justified at all in believing that there are other minds besides your own. You know that you have a mind because you directly experience your own thoughts, perceptions, and emotions, but you have no direct experience of the mental states of others. So how are you justified in believing that other people have mental states?

The second, more moderate problem of other minds assumes that we can have knowledge about the minds of others, but asks how we manage to do this. What kinds of

reasoning or experience enable us indirectly to know what is going on in the minds of people around us? This moderate problem merges with the psychological problem of the nature of the cognitive processes that enable people to understand each other: it belongs as much to the branch of social psychology called social cognition as it does to the philosophy of mind. This section argues that mirror neurons help with both the skeptical and the social-cognition problems of other minds.

Traditional philosophical solutions to the skeptical problem attempt to identify the kinds of inference that justify belief in other minds (Hyslop, 2005). John Stuart Mill and others have construed the inference to other minds as analogical, with a structure such as the following:

> I know by introspection that I have beliefs, desires, and emotions that produce my range of behaviors.
>
> Other people are very similar to me in that they have the same range of verbal and physical behaviors.
>
> Therefore, by analogy, other people also have mental states.

Unfortunately, this argument is rather weak, since it does not deal with the fundamental difference between myself and others, namely that I directly experience my own mental states but simply cannot do so for others. The argument also fails to take into account possible alternative explanations of why other people behave like me, for example that they are robots or zombies or remotely controlled, simulating mental states without actually having them.

More plausibly, belief in the existence of other minds can be seen as justified by a kind of reasoning called *inference to the best explanation* (Graham, 1998). This kind of

inference, which is common in science as well as everyday life, consists in accepting hypotheses if they provide a better explanation of all the available evidence than competing hypotheses. For example, a detective might conclude that one suspect is a murdered because that hypothesis is the best explanation of the forensic evidence such as the presence of the suspect's fingerprints on the murder weapon. For the problem of other minds, the inference runs as follows:

Other people have a wide range of verbal and physical behaviors.

One hypothesis that would explain their behaviors is that they have mental states.

This hypothesis provides a better explanation of the evidence that alternative hypotheses such as that they are robots or zombies.

Therefore, by inference to the best explanation, other people have minds.

The central problem with this inference is: What makes the hypothesis of other minds a better explanation than the alternatives?

According to the account of inference to the best explanation and explanatory coherence that I have developed elsewhere, the most important criteria for evaluating explanations are breadth (how much is explained), simplicity (how little is assumed), and analogy (how similar the explanations are to accepted ones) (see Thagard, 1988, 1992). The hypothesis that there are other minds explains the full range of people's behavior without making a lot of special assumptions that would be required to say why people seem to have mental states when they are really robots or zombies. For example, the hypothesis that other people are robots remotely controlled by aliens requires assumptions about the existence and abilities of the controllers.

Moreover, the inference to other minds incorporates the analogical considerations mentioned earlier, because its overall plausibility is supported by the large number of similarities between my behaviors and those of others. The analogy is enriched by considering not only behavioral but also anatomical and physiological similarity. We know from neuroscience that most people's brains are very similar in both anatomy and functioning. If I wonder about whether you have a mind like mine and I make the reasonable inference that my mental states are closely tied to my brain states, then I can check out the similarities between you and me by having both of us undergo brain scans, with the predictable result that your brain is functioning like mine. Such findings strengthen further the analogy between myself and others, and contribute to the overall coherence and acceptability of my conclusion that you and other people have minds like mine. Just as I have brain states that produce my mental states and behaviors, so do you, by inference to the best explanation.

I think this solution to the skeptical problem of other minds is adequate, but it makes the problem harder than it needs to be if we understand how mirror neurons work. The best-explanation account requires us to have an abstract, verbal, conceptual representation of the mental states of others, opening up the possibility that this representation might be wrong. But mirror neurons provide a person with much more direct understanding of what is going on in the mind of another person. When I see you pick up an apple, I do not need to construct a complex inference about what you are experiencing when you grasp it. Rather, my mirror neurons provide me with a mind-brain process that overlaps with the mind-brain process that occurs when I myself grasp something. Hence I do not have to construct an elaborate inference to understand your

behavior: I just get it. Similarly, when I see you are pricked with a needle I do not need to infer by an analogy-based inference to the best explanation that you are feeling the same kind of pain that I would feel in a similar situation. Instead, my perception of your situation generates an internal brain process in cortical areas such as the insula and anterior cingulate which overlaps with the brain process that produces my own pain. I do not have to infer your pain – I actually feel something like your pain.

Mirror neurons are not by themselves a full solution to the problem of other minds, as there are many mental states, for example whatever you are thinking about this paper right now, that are not directly mirrored. But they do provide a useful supplement to the inference to the best explanation by providing some physically direct connections that do not require complex inference. Similarly, although the philosophical problem of the existence of the external world is arguably solvable by an inference to the best explanation that there is a physical world that explains my sense experiences, we often do not have to make such inferences because of direct causal processes linking world and perception. In vision, for example, we know that light reflects off objects into our eyes, stimulating the neurons in the retina to send signals to layers of brain areas in the visual cortex. Our knowledge of the external world is based as much on this kind of direct causal process as it is on the verbal inference that there is a world independent of the senses. Similarly, in mathematics our understanding of number is not a purely verbal construction, but rests on evolved neural processes for counting and comparing magnitudes that are also found in rats and pigeons (Dehaene, 1997). Just as mathematics grows out of our primitive number sense rooted in neurophysiology, and just as knowledge of the external world grows out of our perceptual ability to interact physically

with the world, so our knowledge of other minds is partly underpinned by a kind of direct understanding via mirror neurons. I will return to the question of what this understanding amounts to at the end of this section.

Just as mirror neurons contribute to a solution to the skeptical problem of other minds, they even more manifestly contribute to a solution to the social-cognition problem of how people manage, albeit imperfectly, to understand the minds of others, a process sometimes called *mentalizing* (Frith and Frith, 1999) or *mindreading* (Stich and Nichols, 2003). There are two main accounts of the nature of mentalizing, as theorizing and as simulating. On the theorizing account (sometimes called the theory theory), people understand others by having explicit causal accounts of why they behave as they do (Gopnik and Meltzoff, 1997). In contrast, on the simulation account, people often understand others by putting themselves in their place and running their own mental states to generate similar experiences and behaviors (Goldman, 2006).

I do not see these as competing theories, but rather as complementary accounts of different ways that people can understand each other (Barnes and Thagard, 1997). Under different circumstances, with different kinds of knowledge available, we will sometimes theorize and at other times simulate, and may in combine these activities by using our theories to improve our simulations and vice versa. Neural mirroring is best viewed as a particularly direct non-verbal kind of simulation (Goldman, 2005; Gallese and Goldman, 1998).

Thus, contrary to the suggestion of Gallese and Rizzolatti (2004), mirror neurons do not provide a novel stand-alone alternative theory of mindreading, but rather a useful supplement to simulation accounts which are in turn a useful complement to the more

easily grasped theory-based account of how we understand other people. Here is an integrated example. Suppose you see your friend John furiously pounding his fist on his desk and cursing. You can understand his action theoretically by applying some kind of general rule or schema that people who are upset often behave in this way, or understand it via simulation by imagining yourself in John's situation (perhaps he just had a paper rejected) and simulating your physical reaction to the bad news, or both. More directly, if there are mirror neurons for anger as there are for disgust and pain, you will understand his situation by virtue of the fact that some of the neurons firing in your brain are the ones that fire when you are angry yourself. Hence the mirror-neuron account provides a useful supplement to theory-based and simulation-based accounts of how we understand other minds.

But does the activation of mirror neurons actually provide understanding? I think it does, but appreciating how requires an extension to familiar conceptions of explanation. Many philosophers and other theorists have maintained that explanations provide understanding of phenomena by locating them in causal networks (Salmon, 1984; Pearl 1988, 2000; Glymour, 2001; Thagard, 1999; Sloman, 2005). For example, we explain why someone got influenza by identifying all the causal factors – genetic makeup, environment, viral infection – that produced fever and the other symptoms of the disease.

In scientific discussions and in ordinary discourse, we normally use verbal descriptions to indicate causal relations, as in the theoretical account that John's paper rejection caused him to be angry which caused him to bang on his desk. I conjecture, however, that our fundamental understanding of causality is not verbal or mathematical.

We know that A causes B is not just a matter of the probability of B given A being higher than the probability of B given not-A, for there may be other interacting factors responsible for the increased probability. Drowning is more common on days when ice cream consumption is high, but that is not because ice cream causes drowning or drowning makes people want ice cream, but rather because hot weather leads to both more drowning and ice cream eating. Pearl (2000) and Woodward (2004) maintain that our understanding of causality is based on the idea of intervention, that A causes B is not just a probabilistic relation but depends on understanding that doing A produces B.

I conjecture that the idea of intervention is essentially visual-motor rather than verbal. Every infant in a crib quickly learns that batting at a mobile or pushing a blanket will create a change: motor behavior produces a visual change. Perhaps there is an innate visual-motor schema for this kind of causal intervention, or perhaps the human brain is just set up to acquire this schema quickly from early experience. Either way, the basic understanding of causality operates well before a child manages to acquire the verbal repertoire to talk about causes, let alone to reason about them using probability theory.

If the causal relation is fundamentally visual-motor, then so must be representations of causes and their effects. The causal schema might be something like this:

<visual-motor representation of action> -->

<visual-motor representation of change>.

From the perspective of neuroscience, there is nothing mysterious about such nonverbal representations. Populations of neurons can encode aspects of the environment in many

ways: visual, motor, other sensory, emotional, as well as verbal (see Eliasmith and Anderson, 2003, for a theory of neural representation). Now we can see how mirror neurons provide understanding, by instantiating visual-motor representations of causal relations. When I observe you being pricked with a pin, it might activate my verbal pain schema: If an object pierces my skin, then I hurt. But in addition mirror neurons provide activation of a nonverbal schema in which the action and its result are both represented nonverbally, by populations of neurons that encode piercing and hurting through a combination of visual, motor, and perceptual correlations. In this way I understand the causes of your behavior by a non-verbal representation of causality.

Hence language and mirror neurons can both provide understanding by locating an occurrence in a causal network, but they do so in different ways because they use different kinds of representations of effects and causes and the relations between them. Given that monkeys and infants seem to have some comprehension of causality, I would argue that the nonverbal kinds of representations – visual, sensory, motor, etc. – are more fundamental to understanding causal relations than our verbal descriptions. It is in this sense that the understanding of other people via mirror neurons is more direct than our verbal versions of why people do what they do. This directness helps both with the skeptical problem of other minds and the social-cognition problem. In the next two sections, I will argue that it also helps with understanding empathy and moral motivation.

EMPATHY

Empathy, where you imagine yourself in someone's situation and get some indication of their emotional state, is important for understanding other people and for making moral decisions about them. Barnes and Thagard (1997) presented an account

of empathy as a kind of analogical mapping relying largely on verbal representations of someone's situation. Mirror neurons make possible a more direct kind of empathy employing visual-motor representations.

Barnes and Thagard (1997) schematize empathy in accord with the cognitive theory of emotions, according to which emotions are primarily indications of the extent to which personal goals are or are not being achieved (Oatley, 1992). For example, people are happy when their goals are being satisfied, and angry when someone or something is blocking the satisfaction of their goals. From this perspective, empathy has the following structure:

The person **P** is in situation **S**, which is like your situation **S'**.

P has goals **G** which are like your goals **G'**.

When you faced situation **S'** which affected your goals **G'**, you felt emotion **E'**. (**S'** and **G'** caused **E'**.)

So **P** is probably feeling emotion **E**, which is like your **E'**, caused by **S** and **G**.

For example, if you want to understand a friend who seems sad because of a disappointment, you can remember a situation such as a job rejection where you were sad because your career goals were not accomplished. Empathy is thus analogical mapping between someone else's situation and your own, and can be easily modeled computationally by ACME, a program which maps between similar situations and transfers information from one to the other (Holyoak and Thagard, 1989). A somewhat richer account of empathy as analogical mapping is accomplished by the model HOTCO, which incorporates ACME but allows the transfer of nonverbal representations of

emotional valences corresponding to positive and negative emotions (Thagard and Shelley, 2001). Still, HOTCO depends on highly verbal representations of the situations of both the provider and recipient of empathy.

In contrast, Singer et al. (2006) advocate a perception-action model of empathy, in which observation or imagination of another person in a particular emotional state automatically activates a representation of that state in the observer (see also Preston and de Waal, 2002; Decety and Jackson, 2004; Jackson et al., 2006; Singer et al., 2004). Singer and her colleagues used functional magnetic resonance imaging (fMRI) to measure brain activity in volunteers who observed others receiving painful stimulation to their hands. As expected, mere observation of another's pain produced increased activation in the pain network of the observer, including the insula and anterior cingulate. As in the earlier study of Singer et al. (2004), people who scored higher on standard empathy scales had higher activity in these brain areas. It thus appears that more empathetic people have more active mirror neuron systems for appreciating the pain of others.

The major manipulation of the Singer et al. (2006) study was that the people who received painful stimulation had previously engaged in a game where some had behaved fairly and others unfairly. Men, and to a lesser extent women, showed much less pain-related brain activation for those sufferers who had acted unfairly. Moreover, men but not women showed greater activation in the reward-related area of the nucleus accumbens when observing unfair people being punished. Thus men more than women took pleasure in the pain of wrongdoers.

The studies of Singer and her colleagues (2004, 2006) suggest a need to expand the largely verbal account of empathetic analogical mapping provided by Barnes and Thagard (1997). The source analog is my own experience of what I experienced as the result of stimulation:

<visual/tactile representation of my stimulation> -->

<sensory/affective representation of my pain>

Then, when I see you stimulated similarly, the result is:

<visual representation of your stimulation> -->

<sensory/affective representation of my AND your pain>.

This mental operation is still a sort of analogical inference, in that it involves grasping a relational similarity between two situations. But it is much more direct than the verbal sort performed by ACME. The arrows indicate causality, but in a way that should not be understood verbally. Rather, as suggested in the last section, causality itself is understood via visual-motor neural representations. Thus feeling your pain can sometimes be a direct reaction based on observation, not an intellectual exercise performed by seeing systematic mappings between two people's situations and goals. The intellectual, verbal kind of empathy may still occur, but it probably depends on the visual/motor/sensory/affective neural pathways that generate the emotional response that is the hallmark of empathy. I may feel your pain as the result of thinking about your situation and seeing parallels with my own experiences, but observing you in pain much more immediately gives me a sense of your pain.

Empathy as verbal analogy fits well with the cognitive theories of emotion advocated by Oatley (1992), Nussbaum (2001), and others; whereas empathy as physical

experiences fits better with the physiological theories of emotion advocated by James (1884) and Damasio (1994). But a full theory of emotion should incorporate both cognitive and physiological processes, and so should an account of the full range of empathy. Hodges and Wegner (1997) review two complementary forms of empathy, one automatic and largely unconscious and the other controlled by conscious inferences.

My mirror-neuron account makes it clear how even the physically direct kind of empathy differs from emotional contagion, which involves picking up an emotion from someone else without any inference. According to Hatfield, Cacioppo, and Rapson (1994, pp. 10-11) theory of emotional contagion, one person "catches" another's emotions as follows:

> 1. In conversation, people tend automatically and continuously to mimic and synchronize their movements with the facial expressions, voices, postures, movements, and instrumental behavior of others.
>
> 2. Subjective emotional experiences are affected, moment to moment, by the activation and/or feedback from such mimicry.
>
> 3. Given propositions 1 and 2, people tend to "catch" others' emotions, moment to moment.

In contrast, empathy via mirror neurons does not require mimicry or behavioral synchronization, but only the perception of another's situation, which activates a kind of perceptual/motor schema that generates an analogous feeling.

In sum, empathy can be based on the kind of verbal analogical mapping discussed by Barnes and Thagard (1997), but more fundamentally can involve direct perceptual detection of the relation between someone's situation and your own via your mirror

neurons. Either way, the phrase "I feel your pain" is not just a touchy-feely cliché, but rather an expression of genuine appreciation of the experiences of others. Moreover, feeling the pain of others can contribute enormously to caring about them and being motivated to act ethically in general. Empathy is a major factor in the moral development of children (Hoffman, 2000).

MORAL MOTIVATION

Why be moral? This question is fundamental for ethics, because even if people can figure out what are the right things to do, we can ask why they would in fact do those things. The problem of moral motivation – what makes people do what is right – has two classes of answers, rationalist and sentimentalist (Nichols, 2004). The traditional philosophical responses to the problem have been rationalist: We should be moral because it would be irrational to do otherwise. The rationality of morality might derive from a priori truths about what is right, as in Kantian deliberation, or from contractarian arguments that it is rational for people to agree with others to be moral, Nichols (2004) argues that a major problem for rationalism is that there is a class of people, psychopaths, who have no impediments in rationality but nevertheless see nothing wrong in harming other people.

Nichols argues convincingly that what is wrong with psychopaths is not their rationality but their emotions. Blair, Mitchell, and Blair (2005) review substantial evidence that psychopathy, whose symptoms include antisocial behavior, lack of guilt, and poverty of emotions, is the result of impairments to emotional learning that derive from disrupted functioning of the amygdala, an area of the brain well known to be important for processing emotions such as fear. That psychopathy derives from

emotional problems fits well with the metaethical position of sentimentalism, according to which moral judgment is grounded in affective response. This tradition goes back to eighteenth-century writers such as David Hume and Adam Smith, and continues today (e.g. Gibbard, 1990; Nichols, 2004; Prinz, 2006).

According to Nichols (2004, p. 98), an adequate sentimalist account must explain how emotion plays a role in linking moral judgment to motivation, while also allowing a place for reason in moral judgment. His explanation is cultural and historical: "Norms are more likely to be preserved in the culture if the norms resonate with our affective systems by prohibiting actions that are likely to elicit negative affect." (Nichols, 2004, p. 140). Norms that prohibit harm to others are virtually ubiquitous across cultures because of this "affective resonance." The adoption of norms makes it possible to reason about what is right and wrong, but these norms have an emotional underpinning that intrinsically provides a connection between morality and action: people are moral because of their emotional commitment to normative rules.

What is missing from Nichols' otherwise plausible account is an explanation of *why* people have such a basic emotional reaction to harm to others. There is no mystery concerning why you do not want harm to yourself, because experiences such as pain and fear are intrinsically negative. Appreciating harm to others might be done by the sort of abstract analogical reasoning discussed by Barnes and Thagard (1997), but there is no guarantee that such reasoning will be motivating: I may understand that you experience pain and fear, but why should I care? What makes emotional moral learning work?

As my discussion of empathy indicated, mirror neurons provide the plausible missing link between personal experience and the experience of others. People not only

observe the pain and disgust of others, they experience their own versions of that pain and disgust, as shown by the mirroring activity in cortical regions such as the insula and anterior cingulate. Normal children do not need to be taught moral rules as abstract theological principles "Thou shalt not kill!" or rational ones "Act only in ways that could become universal." Normal children do not need to reason about why harm is bad for other people: they can actually feel that harm is bad. Thus mirror neurons provide motivation not to harm others by virtue of direct understanding of what it is for another to be harmed.

It would be elegant if there were evidence that psychopaths have deficiencies in the functioning of their mirror neurons, but the relevant experiments have not yet been done. There, is however, recent evidence for mirror neuron dysfunction in autism spectrum disorders (Oberman et al., 2005; Theoret et al., 2005). Individuals with autism have difficulties with imitation, language, theory of mind, and empathy, and it is possible that malfunctioning in mirror neuron systems is at least partly responsible. Similarly, it is possible that psychopaths' deficits in emotional learning, attributed by Blair, Mitchell, and Blair (2005) to disrupted functioning of the amygdala, are partly due to mirror neuron malfunctioning. Children who are incapable for genetic or environmental reasons of feeling the pain of others will not be able to become motivated to follow rules that direct them not to harm other people. Blair, Mitchell, and Blair (2005, p. 128) discuss moral socialization in terms of aversive conditioning, as when caregivers punish children for their wrongdoings. They claim that the sadness, fearfulness, and distress of a victim act as a stimulus to instrumental learning not to produce harm. The involvement

of mirror neurons shows why instrumental learning can be especially effective when people can fully appreciate what is negative about their behavior.

I have argued that mirror neural mechanisms contribute to solution of the philosophical problem of moral motivation by showing how biologically normal people naturally have at least some understanding and concern for harm to other people. Feeling the pain of others is not the whole story of moral motivation, for there are many cognitive and social additions in the form of rules and expectations that can be built on top of neural mirroring. The motivating reason to be moral is not that just that morality is rational, but rather that feeling the pain of others is biologically part of being human.

CONCLUSION

Hence the functioning of mirror neurons contributes to understanding of other minds, empathy, and moral motivation. This functioning does not by itself constitute a solution to any of these problems, but shows how part of the solution in each case is a direct causal connection that complements the more traditional solutions based on conscious, verbal inferences. Such complementation does not make sense if one views justification as akin to verbal deduction, but fits well with a view of justification as coherence involving parallel satisfaction of multiple constraints, including ones represented non-verbally (Thagard, 2000).

The current discussion is only a preliminary to future work in both neuroscience and philosophy. For neuroscience, it would be desirable to have a much more detailed account of how mirroring works, that is how some groups of neurons respond both to one's own experience and to one's perception of others. Perhaps it will be possible to expand current models of cognitive/affective integration to indicate how one person's

emotional state can also be used to appreciate the emotional state of another (Wagar and Thagard, 2004; Litt, Eliasmith, and Thagard, 2006). Such improved models should be able to give a much deeper understanding of the philosophically important problems of other minds, empathy, and moral motivation.

Acknowledgements. For comments on an earlier draft, I am grateful to Lorraine Besser-Jones, Chris Eliasmith, and Abninder Litt. This research was supported by the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

Avenanti, A., Bueti, D., Galati, G., & Aglioti, S. M. (2005). Transcranial magnetic stimulation highlights the sensorimotor side of empathy for pain. *Nature Neuroscience, 8*(7), 955-960.

Barnes, A., & Thagard, P. (1997). Empathy and analogy. *Dialogue: Canadian Philosophical Review, 36*, 705-720.

Blair, J., Mitchell, D. R., & Blair, K. (2005). *The psychopath: Emotion and the brain*. Malden, MA: Blackwell Pub.

Botvinick, M., Jha, A. P., Bylsma, L. M., Fabian, S. A., Solomon, P. E., & Prkachin, K. M. (2005). Viewing facial expressions of pain engages cortical areas involved in the direct experience of pain. *Neuroimage, 25*(1), 312-319.

Damasio, A. R. (1994). *Descartes' error*. New York: G. P. Putnam's Sons.

Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy. *Behav Cogn Neurosci Rev, 3*(2), 71-100.

Dehaene, S. (1997). *The number sense*. New York: Oxford University Press.

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research, 91*(1), 176-180.

Eliasmith, C., & Anderson, C. H. (2003). *Neural engineering: Computation, representation and dynamics in neurobiological systems*. Cambridge, MA: MIT Press.

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

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain, 119 (Pt 2)*, 593-609.

Gallese, V., & Goldman, A. I. (1998). Mirror neurons and the simulation theory of mindreading. *Trends in Cognitive Sciences, 2*, 493-501.

Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Sciences, 8*(9), 396-403.

Gibbard, A. (1990). *Wise choices, apt feelings*. Cambridge, MA: Harvard University Press.

Glymour, C. (2001). *The mind's arrows: Bayes nets and graphical causal models in psychology*. Cambridge, MA: MIT Press.

Goldman, A. I. (2005). *Mirror systems, social understanding and social cognition*. Retrieved May 17, 2006, from http://interdisciplines.org/mirror/papers/3

Goldman, A. I. (2006). *The simulating mind*. New York: Oxford University Press.

Gopnik, A., & Meltzoff, A. N. (1997). *Words, thoughts, and theories*. Cambridge, MA: MIT Press.

Graham, G. (1998). *Philosophy of mind: An introduction* (2nd ed.). Oxford: Blackwell.

Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1994). *Emotional contagion*. Cambridge: Cambridge University Press.

Hodges, S. D., & Wegner, D. M. (1997). Automatic and controlled empathy. In W. Ickes (Ed.), *Empathic accuracy* (pp. 311-339). New York: The Guilford Press.

Hoffman, M. L. (2000). *Empathy and moral development: Implications for caring and justice*. Cambridge: Cambridge University Press.
Holyoak, K. J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science, 13*, 295-355.
Hyslop, A. (2005). *Other minds*. Retrieved May 17, 2006, from http://plato.stanford.edu/archives/win2005/entries/other-minds
Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biology, 3*(3), e79.
Jackson, P. L., Brunet, E., Meltzoff, A. N., & Decety, J. (2006). Empathy examined through the neural mechanisms involved in imagining how I feel versus how you feel pain. *Neuropsychologia, 44*(5), 752-761.
James, W. (1884). What is an emotion? *Mind, 9*, 188-205.
Kohler, E., Keysers, C., Umilta, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: action representation in mirror neurons. *Science, 297*(5582), 846-848.
Litt, A., Eliasmith, C., & Thagard, P. (2006). Why losses loom larger than gains: Modeling neural mechanisms of cognitive-affective interaction. In R. Sun & N. Miyake (Eds.), *Proceedings of the twenty-eighth annual meeting of the Cognitive Science Society* (pp. 495-500). Mahwah, NJ: Erlbaum.
Nelissen, K., Luppino, G., Vanduffel, W., Rizzolatti, G., & Orban, G. A. (2005). Observing others: multiple action representation in the frontal lobe. *Science, 310*(5746), 332-336.
Nichols, S. (2004). *Sentimental rules : on the natural foundations of moral judgment*. Oxford: Oxford University Press.
Nichols, S., & Stich, S. (2003). *Mindreading: An integrated account of pretense, self-awareness, and understanding other minds*. Oxford: Oxford University Press.
Nussbaum, M. (2001). *Upheavals of thought*. Cambridge: Cambridge University Press.
Oatley, K. (1992). *Best laid schemes: The psychology of emotions*. Cambridge: Cambridge University Press.
Oberman, L. M., Hubbard, E. M., McCleery, J. P., Altschuler, E. L., Ramachandran, V. S., & Pineda, J. A. (2005). EEG evidence for mirror neuron dysfunction in autism spectrum disorders. *Brain Res Cogn Brain Res, 24*(2), 190-198.
Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. San Mateo: Morgan Kaufman.
Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge: Cambridge University Press.
Preston, S. D., & de Waal, F. B. (2002). Empathy: Its ultimate and proximate bases. *Behav Brain Sci, 25*(1), 1-20; discussion 20-71.
Prinz, J. (2006). The emotional basis of moral judgments. *Philosophical Explorations, 9*, 29-43.
Rizzolatti, G. (2005). The mirror neuron system and imitation. In S. Hurley & N. Chater (Eds.), *Perspectives on imitation: From neuroscience to social science. Volume 1: Mechanisms of imitation and imitation in animals* (pp. 55-76). Cambridge, MA: MIT Press.

Rizzolatti, G., & Arbib, M. A. (1999). From grasping to speech: imitation might provide a missing link: reply. *Trends in Neurosciences, 22*(4), 152.
Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27*, 169-192.
Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research, 3*(2), 131-141.
Salmon, W. (1984). *Scientific explanation and the causal structure of the world.* Princeton: Princeton University Press.
Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science, 303*(5661), 1157-1162.
Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., & Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature, 439*(7075), 466-469.
Sloman, S. A. (2005). *Causal models: How people think about the world and its alternatives*. New York: Oxford University Press.
Thagard, P. (1988). *Computational philosophy of science*. Cambridge, MA: MIT Press/Bradford Books.
Thagard, P. (1992). *Conceptual revolutions*. Princeton: Princeton University Press.
Thagard, P. (1999). *How scientists explain disease*. Princeton: Princeton University Press.
Thagard, P. (2000). *Coherence in thought and action*. Cambridge, MA: MIT Press.
Thagard, P., & Shelley, C. P. (2001). Emotional analogies and analogical inference. In D. Gentner, K. H. Holyoak & B. K. Kokinov (Eds.), *The analogical mind: Perspectives from cognitive science* (pp. 335-362). Cambridge, MA: MIT Press.
Theoret, H., Halligan, E., Kobayashi, M., Fregni, F., Tager-Flusberg, H., & Pascual-Leone, A. (2005). Impaired motor facilitation during action observation in individuals with autism spectrum disorder. *Current Biology, 15*(3), R84-85.
Wagar, B. M., & Thagard, P. (2004). Spiking Phineas Gage: A neurocomputational theory of cognitive-affective integration in decision making. *Psychological Review, 111*, 67-79.
Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in my insula: The common neural basis of seeing and feeling disgust. *Neuron, 40*(3), 655-664.
Woodward, J. (2004). *Making things happen: A theory of causal explanation*. Oxford: Oxford University Press.