The Neuroeconomics of Mind Reading and Empathy

By TANIA SINGER AND ERNST FEHR*

Economics and game theory are based on the assumption that people are capable of predicting others' actions. The most fundamental solution concepts in game theory (Nash equilibrium, backward induction, and iterated elimination of dominated strategies) are based on this assumption. These concepts require people to be able to view the game from the other players' perspectives (i.e., to understand others' motives and beliefs). Economists still know little about what enables people to put themselves into others' shoes and how this ability interacts with their own preferences and beliefs. In fact, experimental evidence suggests that many people do not obey these concepts and frequently behave as if they believe, counterfactually, that others will play dominated strategies (Colin F. Camerer, 2003). Social neuroscience provides insights into the neural mechanism underlying our capacity to represent others' intentions, beliefs, and desires, referred to as "theory of mind" or "mentalizing," and the capacity to share the feelings of others, referred to as "empathy." We summarize the major findings about the neural basis of mentalizing and empathizing and discuss their implications for economics.

Normal adults are capable of both mentalizing and empathizing. These abilities are useful for making self-interested choices because they enable people to predict others' actions more accurately. However, empathy is also likely to render people less selfish because it allows the sharing of emotions and feelings with others and therefore motivates other-regarding behavior. In fact, neuroscientific empathy experiments indicate that the same affective brain circuits are automatically activated when we feel pain *and* when others feel pain. Therefore, empathy renders our emotions other-regarding, which provides the motivational basis for other-regarding behavior.

I. Mind-reading

For the past several decades, research in developmental psychology, social psychology, and cognitive neuroscience has focused on the human ability to have a "theory of mind" or to "mentalize" (e.g., Uta Frith and Christopher D. Frith, 2003), that is, to make attributions about the mental states (desires, beliefs, intentions) of others. This ability is absent in monkeys and only exists in a rudimentary form in apes (Daniel J. Povinelli and Jess M. Bering, 2002). It develops by about age 5 and is impaired in autism. The lack of a theory of mind in most autistic children could explain their observed failures in communication and social interaction. Recent imaging studies on normal healthy adults have focused on the ability to "mentalize" and have used a wide range of stimuli which represented the intentions, beliefs, and desires of the people involved (for a review, see Helen L. Gallagher and Frith [2003]). Several recent studies, for example, involved the brain imaging of subjects while they played strategic games (Kevin McCabe et al., 2001; Gallagher et al., 2002; Meghana Bhatt and Camerer, 2005) with another partner outside the scanner room. The first two studies examine the brain areas involved when a subject plays against an intentional actor (i.e., another person) as compared to playing against a computer. The study by Bhatt and Camerer explicitly examines brain activity

† *Discussants:* David Laibson, Harvard University; Colin F. Camerer, California Institute of Technology; Kevin McCabe, George Mason University. A third paper, "The Vulcanization of the Human Brain: The Neural Bases of Cognitive-Emotion Interactions in Decision," by Jonathan Cohen, was presented at the meeting session but is not being published in the *Papers and Proceedings.*

* Singer: Wellcome Department of Imaging Neuroscience, 12 Queen Square, WC1N 3BG London, United Kingdom (e-mail: t.singer@fil.ion.ucl.ac.uk); Fehr: Institute for Empirical Research in Economics, Blümlisalpstrasse 10, 8006 Zürich, Switzerland (e-mail: efehr@iew.unizh.ch).

in choice tasks and belief formation tasks. All these studies have repeatedly demonstrated the involvement of one brain area, a part of the medial prefrontal lobe called the anterior paracingulate cortex. This brain area is not only involved when mentalizing about the thoughts, intentions or beliefs of others, but also when people are attending to their own states. Frith and Frith (2003) suggest that this area subserves the formation of decoupled representations of beliefs about the world, "decoupled" in the sense that they are decoupled from the actual state of the world and that they may or may not correspond to reality.

A related line of research has focused on the investigation of the neural mechanism underlying our ability to represent others' goals and intentions by the mere observation of their motor actions. This notion stems from the finding that there are neurons in the premotor cortex of the macaque brain that fire both when the monkey performs a hand action itself and when it merely observes another monkey or a human performing the same hand action (Giacomo Rizzolatti et al., 1996). It has been suggested that these "mirror neurons" represent the neural basis for imitation. Thus, when we imitate someone, we first observe the action and then try to reproduce it. But how do we transform what we see in terms of perceptual input into knowledge of what we need to do in terms of motor commands? The discovery of mirror neurons demonstrated that a translation mechanism is present in the primate brain and automatically elicited when viewing others' actions. Moreover, Vittorio Gallese and Alvin Goldman (1998) suggest that this mirror system might underlie our ability to share others' mental states, providing us with an automatic simulation of their actions, goals, and intentions. A similar common coding of the production and perception of motor action has been demonstrated in the human brain using imaging techniques such as PET and fMRI since the discovery of these "mirror neurons" (for a review, see Julie Grezes and Jean Decety [2001]).

II. Empathy

In addition to the ability to understand mental states of others, humans can also empathize with others, that is, share their feelings and emotions in the absence of any direct emotional stimulation to themselves. Humans can feel empathy for other people in a wide variety of contexts: for basic emotions and sensations such as anger, fear, sadness, joy, pain, and lust, as well as for more complex emotions such as guilt, embarrassment, and love. The idea that a neural system enables people to share others' mental states has recently been expanded to include the ability to share their feelings and sensations (e.g., Stephanie D. Preston and Frans B. M. de Waal, 2002). How can we understand what someone else feels when he or she experiences emotions such as sadness or happiness, or bodily sensations such as pain, touch, or tickling, in the absence of any emotional or sensory stimulation to our own body? Influenced by perception–action models of motor behavior and imitation, Preston and de Waal (2002) proposed a neuroscientific model of empathy, suggesting that observation or imagination of another person in a particular emotional state automatically activates a representation of that state in the observer with its associated autonomic and somatic responses. The term "automatic" in this case refers to a process that does not require conscious and effortful processing, but which can nevertheless be inhibited or controlled.

Imaging studies in the last two years have started to investigate brain activity associated with different empathic responses in the domain of touch, smell, and pain. The results have revealed common neural responses elicited by observation of pictures showing disgusted faces and smelling disgusting odors oneself (Bruno Wicker et al., 2003) and by being touched and observing someone else being touched in a video (Christian Keysers et al., 2004). Another study could identify shared and unique networks involved in empathy for pain (Singer et al., 2004b). We will explain the latter study in more detail in order to illustrate how empathic responses can be measured using functional MRI.

In this study, couples who were in love with each other were recruited; empathy was assessed "in vivo" by bringing both woman and man into the same scanner environment. More specifically, brain activity was assessed in the female partner while painful stimulation was applied either to her own or to her partner's

right hand via electrodes attached to the back of the hand. The male partner was seated next to the MRI scanner, and a mirror system allowed the female partner to see both her own and her partners' hands lying on a tilted board in front of her. Flashes of different colors on a big screen behind the board pointed either to her hand or that of her partner, indicating which of them would receive the painful stimulation and which would be subject to the non-painful stimulation. This procedure enabled the measurement of pain-related brain activation when pain was applied to the scanned subject (the so-called "pain matrix") or to her partner (empathy for pain).

The results suggest that some parts, but not the entire "pain matrix," were activated when empathizing with the pain of others. Activity in the primary and secondary somato-sensory cortex was only observed when the subject was receiving pain. These areas are known to be involved in the processing of the sensory-discriminatory components of our pain experience; that is, they indicate the location of the pain and its objective quality. In contrast, bilateral anterior insula (AI), the rostral anterior cingulate cortex (ACC), brainstem, and cerebellum were activated when subjects either received pain or a signal that a loved one experienced pain. These areas are involved in the processing of the affective component of pain, that is, how unpleasant the subjectively felt pain is. Thus, both the experience of pain to oneself and the knowledge that a loved partner experiences pain activates the same affective pain circuits, suggesting that if a loved partner suffers pain, our brains also make us suffer from this pain. These findings suggest that we use representations reflecting our own emotional responses to pain to understand how the pain of others feels. Moreover, our ability to empathize may have evolved from a system which represents our own internal feeling states and allows us to predict the affective outcomes of an event for ourselves and for other people.

The results of the Singer et al. (2004b) study further suggest that the empathic response is rather automatic and does not require active engagement of some explicit judgments about others' feelings. The scanned subject did not know that the experiment was about empathy; subjects were just instructed to do nothing but observe the flashes that indicate either pain to the subject or the loved partner. The analysis also confirmed that the ability to empathize is heterogeneous across individuals; standard empathy questionnaires and the strength of the activation in the affective pain regions (AI and ACC) when the partner received pain were used to assess this heterogeneity. Interestingly, individual heterogeneity measured by the empathy questionnaire was highly correlated with individual differences that were measured by brain activation in AI and ACC. Thus, neural evidence and questionnaire evidence on empathy reinforce each other.

Does empathy also extend to unknown persons? The results of three recent studies indicate that empathic responses are also elicited when scanned subjects do not know the person in pain. Activity in ACC and AI has also been observed when subjects witness still pictures depicting body parts involved in possibly painful situations (Philip L. Jackson et al., 2005) or videos showing a needle stinging in the back of a hand (India Morrison et al., 2004). At the moment, Singer and collaborators are investigating whether the level of empathic response in ACC and AI can be modulated by whether the subject likes or dislikes the "object of empathy." In this study, actors are paid to pretend to be naive subjects participating in two independent experiments, one on "social exchange" the other one on the "processing of pain."

In the first experiment, the two confederates repeatedly play a sequential Prisoner's Dilemma game in the position of the second mover with the scanned subject. One actor plays a fair strategy and usually reciprocates cooperative first-mover choices with cooperation; the other actor plays unfairly and defects in response to first-mover cooperation most of the time. Based on behavioral and neuronal findings of a previous imaging study which revealed verbally reported liking and disliking as well as emotion-related brain activation in responses to faces of people who had previously cooperated or defected (Singer et al., 2004a), we expect to induce subjects to like fair players and to dislike unfair ones.

In the second part of the experiment, all three players participate in a pain study that expands the approach by Singer et al. (2004b). One actor sits on each side of the scanner, enabling the

scanned subject to observe flashes of different colors indicating high or low pain stimulation to his/her hand or to those of the fair or unfair players. We predict empathy-related activation in ACC and AI when observing the unfamiliar but likeable person receiving painful stimulation. However, based on the results of a recent imaging study that reports reward-related activity when players could punish defectors in a sequential Prisoner's Dilemma game (Dominique DeQuervain et al., 2004), we further predict a lack of empathy-related brain activation and an increase in activity in reward-related areas when perceiving a previous defector getting pain, that is, getting punished. Such a pattern of results would contribute to the microfoundation for theories of social preferences. These theories suggest that people's valuations of other players' payoffs depend on the fairness of their previous behavior (Fehr and Simon Gächter, 2000): many people value others' payoffs positively if others behaved fairly; however, people also value others' payoffs negatively if they behaved unfairly. This pattern of preferences implies that people prefer cooperating with fair opponents while favoring the punishment of unfair opponents.

III. Implications for Economics

Mind-reading and empathy are two lines of research which have recently emerged in social neuroscience. Even though these abilities seem to rely on different neural circuitries, the concepts do in fact have common features. Both allow humans to represent states of other people: others' intentions, beliefs, and thoughts or their feeling states based on emotions and sensations. These abilities enable people to predict others' behavior and, therefore, help them meet their individual goals. As an example, imagine that you are a first mover in a social exchange situation like the sequential Prisoner's Dilemma. Your attempt to predict whether the opponent will reciprocate a cooperative choice will rely on your belief about his type (i.e., whether you believe him to be a fair person with a desire to reciprocate or not). However, if you believe that the other person is a reciprocator, you also need to understand his *actual* feeling and motivational state. If, for example, the other player is angry because you repeatedly violated his sense of fairness, he will probably not reciprocate your trust. Your capacity to empathize, that is, to simulate the internal state resulting from being cheated in a social exchange will help you to predict the opponent's likely action. Thus, the ability to empathize is useful from a self-interested point of view. However, the very ability to empathize may also undermine purely self-interested choices and may promote other-regarding behavior. In fact, there is evidence (Nancy Eisenberg and P. A. Miller, 1987) suggesting that affective concern for others and perspective taking is positively related to prosocial behavior (defined as voluntary behavior intended to benefit others).

An important feature of the outlined mechanisms is that they mostly rely on automatic processes. We represent the goals of others in terms of our own goals, without even being aware of it. Without thinking, the perceived feelings of others automatically activate brain networks that also represent our own feeling states; we automatically share other people's feelings. Thus, as our own feelings and emotions are important determinants of our motives, our behavior may be automatically other-regarding unless we inhibit the other-regarding impulses. Therefore, empathic concern may establish a link between the ability to predict others' motives and the nature of own motives, that is, other people's emotions may partly shape our own motives toward them. To provide an example: if shown a picture of a malnourished child with a swollen belly, many people empathize with this child and are therefore willing to incur cost to help the child (e.g., by donating money to charities that operate in third-world countries).

The study by Singer et al. (2004b) suggests that there are individual differences in empathic abilities. Therefore, the hypothesized link between empathic abilities and the prediction of other players' motives and actions suggests a testable prediction: people with stronger empathic abilities are better predictors of others' motives and actions. Moreover, the hypothesis that empathy enhances other-regarding behavior in combination with the existence of individual differences in empathy suggests that people who exhibit more affective concern are more likely to display altruistic behaviors. In

analogy to the findings of Singer et al. (2004b) we also predict that people with higher scores in their perspective-taking ability should display higher activation in areas shown to be activated by "theory of mind" tasks (e.g., mPFC), and consequently these people should also be better in predicting the actions of others. An interesting question for future research is to determine the relative importance of our ability to empathize and to mentalize for the prediction of motives and actions of others in different situations.

Neuroscientific research on mentalizing and empathizing may also help explain how individuals actually assess other players' types in games with incomplete information about preferences. Economists take a technical shortcut in games with incomplete information by assuming a common prior distribution over players' potential preferences ("types"). While this shortcut has enabled economists to solve games with incomplete information, the question about the determinants of this prior probability distribution has not been addressed. In fact, the assumption of a prior distribution over types constitutes a huge black box. Neuroeconomic research may help us to understand what is going on in this black box.

REFERENCES

Bhatt, Meghana and Camerer, Colin F. "Self-Referential Thinking and Equilibrium as State of Mind in Games: fMRI Evidence." *Games and Economic Behavior*, 2005 (forthcoming).

Camerer, Colin F. *Behavioral game theory—Experiments in strategic interaction*. Princeton, NJ: Princeton University Press, 2003.

DeQuervain, Dominique; Fischbacher, Urs; Treyer, Valerie; Schellhammer, Melanie; Schnyder, Ulrich; Buck, Alfred and Fehr, Ernst. "The Neural Basis of Altruistic Punishment." *Science*, 2004, *305*(5688), pp. 1254–58.

Eisenberg, Nancy and Miller, P. A. "The Relation of Empathy to Prosocial and Related Behaviors." *Psychological Bulletin*, 1987, *101*(1), pp. 91–119.

Fehr, Ernst and Gächter, Simon. "Fairness and Retaliation—The Economics of Reciprocity." *Journal of Economic Perspectives*, 2000, *14*(3), pp. 159–81.

Frith, Uta and Frith, Christopher D. "Development and Neurophysiology of Mentalizing." *Philosophical Transactions of the Royal Society of London, B: Biological Sciences*, 2003, *358*(1431), pp. 459–73.

Gallagher, Helen L.; Jack, Anthony I.; Roepstorff, Andreas and Frith, Christopher D. "Imaging the Intentional Stance in a Competitive Game." *NeuroImage*, 2002, *16*(3), Part 1, pp. 814–21.

Gallagher, Helen L. and Frith, Christopher D. "Functional Imaging of 'Theory of Mind' 5." *Trends in Cognitive Science*, 2003, *7*(2), pp. 77–83.

Gallese, Vittorio and Goldman, Alvin. "Mirror Neurons and the Simulation Theory of Mind-Reading." *Trends in Cognitive Sciences*, 1998, *2*(12), pp. 493–501.

Grezes, Julie and Decety, Jean. "Functional Anatomy of Execution, Mental Simulation, Observation, and Verb Generation of Actions: A Meta-analysis." *Human Brain Mapping*, 2001, *12*(1), pp. 1–19.

Jackson, Philip L.; Meltzoff, Andrew N. and Decety, Jean. "How Do We Perceive the Pain of Others: A Window into the Neural Processes Involved in Empathy." *NeuroImage*, 2005, *24*(3), pp. 771–79.

Keysers, Christian; Wicker, Bruno; Gazzola, Valeria; Anton, Jean-Luc; Fogassi, Leonardo and Gallese, Vittorio. "A Touching Sight: SII/PV Activation During the Observation and Experience of Touch." *Neuron*, 2004, *42*(2), pp. 335–46.

McCabe, Kevin; Houser, Daniel; Ryan, Lee; Smith, Vernon and Trouard, Theodore. "A Functional Imaging Study of Cooperation in Two-Person Reciprocal Exchange." *Proceedings of the National Academy of Sciences (USA)*, 2001, *98*(20), pp. 11832–35.

Morrison, India; Lloyd, Donna; di Pellegrino, Giuseppe and Roberts, Neil. "Vicarious Responses to Pain in Anterior Cingulate Cortex: Is Empathy a Multisensory Issue?" *Cognitive, Affective, and Behavioral Neuroscience*, 2004, *4*(2), pp. 270–78.

Povinelli, Daniel J. and Bering, Jesse M. "The Mentality of Apes Revisited." *Current Directions in Psychological Science*, 2002, *11*(4), pp. 115–19.

Preston, Stephanie D. and de Waal, Frans B. M. "Empathy: Its Ultimate and Proximate

Bases." *Behavioral and Brain Science*, 2002, *25*(1), pp. 1–72.

Rizzolatti, Giacomo; Fadiga, Luciano; Gallese, Vittorio and Fogassi, Leonardo. "Premotor Cortex and the Recognition of Motor Actions." *Cognitive Brain Research*, 1996, *3*(2), pp. 131–41.

Singer, Tania; Kiebel, Stefan J.; Winston, Joel S.; Dolan, Ray J. and Frith, Christopher D. "Brain Responses to the Acquired Moral Status of Faces." *Neuron*, 2004a, *41*(4), pp. 653–62.

Singer, Tania; Seymour, Ben; O'Doherty, John P.; Kaube, Holger; Dolan, Ray J. and Frith, Christopher D. "Empathy for Pain Involves the Affective but not Sensory Components of Pain." *Science*, 2004b, *303*(5661), pp. 1157–62.

Wicker, Bruno; Keysers, Christian; Plailly, Jane; Royet, Jean-Pierre; Gallese, Vittorio and Rizzolatti, Giacomo. "Both of Us Disgusted in My Insula: The Common Neural Basis of Seeing and Feeling Disgust." *Neuron*, 2003, *40*(3), pp. 655–64.