TOPICS
TOPICS IN COGNITIVE SCIENCE

Topics in Cognitive Science 2 (2010) 579–595
Copyright © 2010 Cognitive Science Society, Inc. All rights reserved.
ISSN: 1756-8757 print / 1756-8765 online
DOI: 10.1111/j.1756-8765.2010.01086.x

Toward an Integrated Neuroscience of Morality: The Contribution of Neuroeconomics to Moral Cognition

Trevor Kvaran, Alan G. Sanfey

Department of Psychology, University of Arizona

Received 4 February 2009; received in revised form 5 May 2009; accepted 21 May 2009

Abstract

Interest in the neural processes underlying decision making has led to a flurry of recent research in the fields of both moral psychology and neuroeconomics. In this paper, we first review some important findings from both disciplines, and then argue that the two fields can mutually benefit each other. A more explicit recognition of the role of values and norms will likely lead to more accurate models of decision making for neuroeconomists, whereas the tasks, insights into neural mechanisms, and mathematical modeling common in neuroeconomic research offer moral psychologists the opportunity to expand their field and move beyond methodological limitations that may have hindered the field's progress to this point. We conclude by highlighting an exciting group of recent studies that illustrate the potential of research that embraces the integrated moral/neuroeconomic approach that we suggest here.

Keywords: Decision making; Neuroeconomics; Moral judgment; Social neuroscience

1. Introduction

Morality is an inescapable aspect of life. Nearly all cultures have detailed moral codes, many of them remarkably similar, and from an early age children are able to both appreciate moral rules and incorporate them into their behavior (Piaget, 1965/1932). However, despite playing such a prominent role in everyday life, the scientific study of morality has only recently emerged as a vibrant field of research. Traditionally, ethical concerns have been the remit of philosophers and theologians, who have focused on normative questions of how people *ought* to act. Buttressing these normative approaches, psychologists and neuroscientists have in recent years begun to explore the mechanisms that may underlie moral

Correspondence should be sent to Alan G. Sanfey, Department of Psychology, University of Arizona, 1503 E University Blvd., Tucson, AZ 85721. E-mail: asanfey@u.arizona.edu

judgments and decisions by investigating processing at both the cognitive and neural levels. In a relatively short period of time, this research has greatly expanded our understanding of the moral mind and led to new ways of considering how moral judgments and decisions are made. This promising new research direction utilizes a largely interdisciplinary approach, merging philosophy, psychology, and neuroscience toward the goal of understanding the psychology of morality.

In concert, neuroeconomics, another interdisciplinary field, has also recently begun to study the neural substrates underlying decision making, in this case typically using economic situations where we are attempting to maximize our own financial gain. In a similar vein to the study of moral decision making, neuroeconomics has already made important contributions to our understanding of the economic mind (Glimcher, Camerer, Fehr, & Poldrack, 2008).

Our thesis is that these two fields have much to offer each other. We aim to illustrate here why the integration of neuroeconomics and moral neuroscience has the potential to benefit both fields, and to focus in particular on how approaches currently being used in neuroeconomics have a useful place in the examination of the moral mind. We will begin by first reviewing the most recent literature on moral judgment and then introduce the most relevant concepts and data from the neuroeconomic approach, interspersed with a discussion of how the interaction of the fields could fruitfully take place.

2. Cognitive neuroscience of morality

Research on the psychological and neural foundations of morality have to date used a variety of methods and tasks to address these important questions, which will be reviewed briefly in the next section. Additionally, one important research direction that has emerged from this interdisciplinary investigation has focused on how emotions may play a prominent role in how we assess morality.

2.1. Methods and tasks

The methods used in this new field have tracked those employed by cognitive neuroscience more broadly over the past decade. In addition to the standard methods of experimental psychology, where participants' judgments and reaction times are measured as they are presented with vignettes involving moral issues, the field has also begun to use methodological approaches allowing for inferences about the neural processes underlying moral judgment.

Patients who have suffered from focal brain lesions have long been used in experimental psychology to determine whether specific brain regions are involved in particular cognitive processes. The use of these participants has also informed the study of moral judgment, as will be discussed below (Ciaramelli, Muccioli, Ladavas, & di Pellegrino, 2007; Koenigs et al., 2007). More recently, there has also been the extensive employment of various neuroimaging techniques, primarily functional magnetic resonance imaging (fMRI; e.g., Greene, Sommerville, Nystrom, Darley, & Cohen, 2001; Schaich-Borg, Hynes, Van Horn, Grafton, & Sinnott-Armstrong, 2006). By measuring the ratio of oxygenated to deoxygenated blood in the brain as the participant ponders a moral scenario, this method takes advantage of the close relationship between this oxygenation ratio and neuronal firing, and thus allows researchers to infer what regions of the brain are involved while the participant makes his or her judgment. Although there are still some open questions with regard to how accurately these regional blood flow changes track neuronal firing, this method has the major advantage of being both noninvasive and also allowing for good spatial and temporal resolution of the judgment process. In addition, other techniques such as transcranial magnetic stimulation (TMS), which allows for temporary modulation of neural activity in healthy participants, are beginning to offer promise in studying these processes.

In terms of how moral judgment is elicited, the vast majority of research on moral cognition has used short vignettes to present a particular situation to the participant, with this scenario usually containing unclear or opposing moral principles. Then the participant is asked either to determine what he or she would have done in that situation, or what he or she believes is morally permissible in the scenario. Probably the most famous set of moral dilemmas are the so-called trolley problems (Thompson, 1985). In the basic version of this scenario, participants are told that they are witness to a trolley careering down a track. In the ''switch'' version of the trolley problem, the participant is standing next to a switch for a railroad. If left on its current course, the trolley will hit and kill five people trapped further down the track. If the participant pulls the switch, the train will be diverted onto an alternate track where only one individual is trapped. Thus, pulling the switch will save five, but will kill one. Participants are usually asked whether it is morally permissible to pull the switch. In the ''footbridge'' version of the scenario, the participant is now standing on a footbridge over the track when he or she notices the runaway trolley. As in the ''switch'' case, the train is headed toward five men trapped on the track. The participant on the footbridge happens to be standing next to a very large man. They are told that if they push the man onto the tracks, the man will stop the trolley, killing him but saving the five others trapped on the track. What is morally permissible here?

A variety of different studies have found that most people say that it is morally permissible to pull the switch, but it is not permissible to push the man off the footbridge, despite the identical number of potential casualties in both cases (1 life vs. 5 lives) (Bartels, 2008; Greene, Morelli, Lowenberg, Nystrom, & Cohen, 2008; Greene, Nystrom, Engell, Darley, & Cohen, 2004; Greene et al., 2001; Hauser, Cushman, Young, Jin, & Mikhail, 2007; Koenigs et al., 2007; Petrinovich, O'Neill, & Jorgensen, 1993; Schaich-Borg et al., 2006; Valdesolo & DeSteno, 2006). The trolley problem has been highly successful in demonstrating that even though surface details of scenarios can be identical, moral judgment is more complicated than merely completing an actuarial accounting of the outcome. However, as will be discussed later, use of these rather unrealistic vignettes may compromise the ability of the participant to really place himself or herself in the situation, leaving open the question of the extent to which theories of moral judgment constructed from research on moral dilemmas will generalize to behavior in the real world.

2.2. Emotion in moral judgment

Perhaps the single most important contribution the last decade of research has made to the field of moral psychology is the notion that emotions play a critical role in moral judgment and decision making. Although by no means a new idea (e.g., see Hume 1739/1978 or Smith 1759/1966, for classical conceptions of the role of sentiments in morality), recent evidence suggesting that emotion is crucial for normal moral judgment directly contradicted the then-dominant Kohlbergian view. Lawrence Kohlberg, building from Piaget's work on the stages of cognitive development (Piaget, 1965/1932), posited that moral judgments are the result of conscious reasoning via moral rules that are developed across the life span (Kohlberg, 1981). His focus on reasoning is a hallmark of rationalist moral psychology, which dominated the field through the latter part of the 20th century, and this perspective made little room for the role of emotion in moral psychology. However, the last decade has seen multiple studies—across different labs, using varying methodologies and techniques—which provide evidence that emotions are crucially involved in the formation of moral judgments.

Much of this research has utilized modern neuroscientific techniques, particularly fMRI, and studies of patients with brain damage (Ciaramelli et al., 2007; Greene et al., 2001, 2004; Koenigs et al., 2007; Moll, de Oliveira-Souza, Bramati, & Grafman, 2002; Moll, de Oliveira-Souza, Eslinger, Bramati, & Mourão-Miranda, 2002; Schaich-Borg et al., 2006). Other researchers have used subtle mood induction primes to investigate the role of particular emotions, such as disgust (Haidt, 2001; Schnall, Haidt, Clore, & Jordan, 2008; Wheatley & Haidt, 2005) and happiness (Valdesolo & DeSteno, 2006). The picture that arises from this body of research is one in which emotions are critical to conceptualizing and implementing morality.

The initial reevaluation of the importance of emotion to moral psychology developed from research that examined the behavioral deficits of patients who had suffered focal brain damage to the prefrontal cortex, particularly the orbitofrontal cortex and ventromedial prefrontal cortex (VMPFC). These studies demonstrated decision-making and emotional deficits, including diminished empathy and increased reckless and antisocial behavior (Anderson, Bechara, Damasion, Tranel, & Damasio, 1999; Damasio, 1994). The predictable patterns of emotional deficits and increased antisocial behavior led to the hypothesis that, counter to the Kohlbergian view, emotional processing may indeed be involved in moral psychology.

In an early attempt at exploring the brain mechanisms underlying morality, Moll and colleagues examined brain activity while participants read and made silent judgments about sentences that either did or did not contain morally relevant information (De Oliveira-Souza & Moll, 2000; Moll, Eslinger, & de Oliveira-Souza, 2001; Moll, de Oliveira-Souza, Bramati, et al., 2002). In each of these studies, increased activation in VMPFC and the medial frontal gyrus was found in moral when compared with nonmoral cases. Employing a different approach, Moll, de Oliveira-Souza, Eslinger, et al. (2002) presented participants in an MRI scanner with emotionally charged images that either contained moral content (i.e., physical assaults or images of war) or did not (i.e., body lesions or dangerous animals).

They found that, as expected, a complex series of brain regions were activated when looking at either moral or nonmoral emotional images. However, they found that activation in the VMPFC and the medial frontal gyrus was selective for the morally charged images. This study was consistent with the early work with brain-damaged patients. Greene et al. (2001, 2004) have also explored emotional factors in moral judgment in a series of well-known functional imaging studies in which participants provide judgments regarding a series of complex moral dilemmas. Greene has distinguished not merely between moral and nonmoral cases but also between different types of moral dilemmas. This program of research makes a distinction between *personal* moral dilemmas, in which the choice faced is ''up close and personal,'' and *impersonal* moral dilemmas, in which the decision maker is more removed from the situation. These two types of moral dilemma are illustrated nicely in the two trolley problems discussed above.

Within Greene's personal⁄impersonal distinction, the ''switch'' case is a prototypical impersonal case, whereas the ''footbridge'' is a prototypical personal case. Greene has suggested that personal cases are likely to induce much more emotional responses than impersonal cases, arguing that this may provide an explanation of the differences in responses we see to the ''switch'' and ''footbridge'' cases. As evidence for the role of emotion in ''personal'' moral dilemmas, Greene et al. (2001) compared neural activity when individuals read and made judgments about both types of dilemma, and found increased activation in the posterior cingulate and angular gyrus for personal compared to impersonal dilemmas. Both of these regions have been associated with emotional processing, bolstering the claim that emotion is involved in personal moral dilemmas.

Taken together, the evidence so far has strongly implicated the VMPFC, an area known to be involved in emotional processing, in moral judgment. Further support for the role of VMPFC in moral judgment comes from a recent study comparing the performance of normal controls to brain-damaged individuals with VMPFC lesions on a moral dilemma task (Koenigs et al., 2007).[1] This study found that, in ''high-conflict'' personal moral dilemmas, brain-damaged individuals were more likely to make the utilitarian choice than were controls (for instance, they were more likely to say that it is permissible to push the large man off the footbridge). This finding was replicated in a similar study with an independent population of VMPFC patients (Ciaramelli et al., 2007). An additional study by Schaich-Borg et al. (2006) found increased activation in brain regions related to emotion, including the VMPFC, when individuals made judgments about cases that involved causing harm in order to ensure an overall benefit.

Although most of the above studies have focused on emotion as a unitary category, several recent studies have also begun to look at specific moral emotions such as guilt, indignation, disgust, and compassion. Two recent studies (Moll et al., 2007; Zahn et al., 2009) found that reading short statements designed to elicit prosocial emotions (guilt, compassion) resulted in increased activity in both anterior VMPFC and the superior temporal sulcus (STS), whereas statements designed to elicit other regarding emotions (disgust, indignation) showed increased amygdala activation. Although research is only beginning to reveal the complicated nature of the neural mechanisms underlying moral emotions, these early studies suggest that moral psychologists should move beyond the simplistic notion of emotion as a unitary concept.

Behavioral studies have also indicated that emotions play an important role in moral judgment. Social Intuitionist theories have proposed that moral judgments are primarily caused by intuitive emotional responses to moral scenarios, and that although conscious reasoning can play a role in judgment formation, it typically serves to generate post-hoc rationalizations for the judgment (Haidt, 2001; Haidt & Bjorklund, 2008). Work on the concept of ''moral dumbfounding'' has been introduced as evidence for this position (Bjorklund, Haidt, & Murphy, 2000). ''Moral dumbfounding'' is the phenomenon in which an individual is highly confident of the rightness or wrongness of an action, but unable to provide a reasonable justification for why that judgment is correct. One well-known example occurs when people are asked to make a moral judgment about the rightness or wrongness of consensual incest (Bjorklund et al., 2000). Most people agree that it is wrong for two adult siblings to engage in incest even when birth control is used and no physical or psychological trauma will occur from the act. When asked why it is wrong, however, most people struggle to provide a reason although they still maintain that the action is wrong. Similar instances of moral dumbfounding have been reported in other labs with substantially different stimuli (Cushman, Young, & Hauser, 2006; Hauser et al., 2007).

Several studies have additionally found that emotional manipulations can significantly impact moral judgment. The emotion that has perhaps received the most attention is disgust. Schnall et al. (2008) have found that individuals make harsher moral judgments in a physically dirty room when compared with a clean room, after smelling a disgusting versus a neutral smell, and after watching a disgusting video when compared with a sad or neutral one. Additionally, participants hypnotically primed to experience disgust judged scenarios as more morally wrong than individuals who had not received hypnotic suggestion (Wheatley & Haidt, 2005).

Although this evidence as a whole appears to leave little doubt that emotion is heavily involved in moral judgment, it is very much an open question exactly what role it plays. Moreover, where does the newfound prominence of emotion leave conscious reasoning in our understanding of moral judgment? One prominent explanation is that moral judgments are the result of a dual-process model (Greene et al., 2001). According to this view, moral dilemmas provoke responses from two separable, and oftentimes competing neural processes, one of which is associated with fast, automatic, affect-laden processing and the other of which is associated with more conscious, deliberate, and controlled reasoning. According to this view, when faced with a dilemma such as the footbridge problem discussed earlier, two competing responses are generated: an affectively aversive response to the idea of pushing the man onto the tracks, combined with a cognitive bias that pushing the man would bring about the greatest overall good in terms of lives saved. This theory proposes that in personal dilemmas such as the footbridge, the aversive response to the thought of pushing the person to his doom overwhelms any concerns about maximizing the overall good, thus generating the nonutilitarian response that it is wrong to push the man. In contrast, the typical utilitarian response to impersonal dilemmas such as the ''switch'' case is explained because the impersonal nature of the scenario causes less of an emotional response, allowing more deliberative concerns about the overall good to be considered.

Much of the evidence on emotions reviewed above is consistent with this position. As already indicated, Greene and colleagues found that personal moral dilemmas were correlated with increased activity in emotion-related brain regions. Additionally, and crucially for their dual-process model, they also found that when individuals read ''impersonal'' moral dilemmas there was significantly increased activation in brain regions associated with working memory and reasoning.

In a follow-up study, Greene et al. (2004) found that in particularly difficult personal moral dilemmas (such as whether to smother an infant to save a group of people), activation was seen in the anterior cingulate cortex, a region thought to be active when cognitive and decision conflict is present (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Pochon, Riis, Sanfey, Nystrom, & Cohen, 2008). This is consistent with the idea that in difficult dilemmas there is significant conflict between the immediate aversive response to harming someone and the utilitarian intuition to maximize the overall good. Moreover, they found that individuals who made the difficult utilitarian choice in personal dilemmas showed increased activation in the dorsolateral prefrontal cortex, a region associated with cognitive control and executive function (Miller & Cohen, 2001). Greene and colleagues interpreted this as support for the dual-process model, arguing that making the utilitarian judgment required increased executive control in order to override the competing prepotent emotional response.

The dual-process model can also accommodate and explain the increased utilitarian responses of VMPFC reported above (Greene, 2007; although see Moll and de Oliveira-Souza, 2007b, for an alternative explanation). According to the dual-process view, damage to brain regions associated with emotional processing should lead to increased utilitarian judgments because there will be less competition from the emotional system. Recently, it has also been shown that putting individuals under increased cognitive load while they engage in a moral dilemma task causes them to be slower to make utilitarian, but not non-utilitarian, judgments in personal moral dilemmas, again consistent with the dual-process model, although it should be noted that this result was restricted to reaction time effects, not differences in moral judgments. Increased cognitive load should interfere with the cognitive processes thought to elicit utilitarian responses, but not with the emotional processes thought to elicit nonutilitarian responses. Valdesolo and DeSteno (2006) provided additional support for the view. They used humorous video clips to induce a positive emotional state and then had people respond to standard trolley problems. They found, consistent with the dual-process model, that people primed with humor were more likely to make the utilitarian response than those who were not. These studies, taken together, suggest that the dual-process model of moral judgment is at the very least a promising theoretical model. However, it should be noted that there is some controversy regarding this proposed model, with some researchers pointing out that there is to date relatively limited evidence for the existence of neural systems that correspond to the purported dual processes (Glimcher, Dorris, & Bayer, 2005).

As this brief review has hopefully made clear, it seems evident that emotion does play a crucial role in moral judgment. Exactly what role it plays though, is still far from clear, and much work is needed in order to fully explain the processes underlying moral judgment. An additional issue in the field of moral judgment is the nature of the tasks used, namely the types of vignettes presented to the participants. As is nicely demonstrated by the trolley series of problems, these scenarios are often rather fantastical (pushing overweight men onto trolley tracks, and so on) and appear to be quite unrepresentative of the type of moral decision making we are faced with in everyday lives.

However, researchers studying different types of decisions have made some useful progress in both understanding how emotions can impact choices and also with the development of tasks in which people are placed in actual consequential situations where moral issues are raised. The following section will briefly review the progress in this field, popularly termed *neuroeconomics*.

3. Neuroeconomics

The emerging field of neuroeconomics, which has applied an interdisciplinary approach to the study of decision making, has the potential to be useful in furthering our understanding of moral psychology. The field makes use of contemporary neuroscience techniques at multiple levels, combined with economic and psychological theories, in order to investigate the neural basis of judgment and decision making. Much like the study of morality, economic behavior has historically been studied from a normative perspective. Research has focused on developing optimal solutions to various types of problems, and of prescribing how one ought to behave when making (typically economic) decisions.

This approach began to change in the 1970s as psychological models of decision making arose that provided a more accurate account of the ways that individuals actually make decisions (Kahneman & Tversky, 1979; Tversky & Kahneman, 1974). The last 30 years have seen a rich body of research aimed at providing psychological explanations for the deviations seen between normative economic models and observed human behavior (Gilovich, Griffin, & Kahneman, 2002; Kahneman, Slovic, & Tversky, 1982; Kahneman & Tversky, 2000). Neuroeconomics has taken this approach one step further by using neuroscientific methods, from single-electrode recordings in primates as they make simple decisions (Dorris & Glimcher, 2004) to noninvasive methods such as fMRI and TMS. We will focus here on studies that have made use of fMRI and TMS, as they are most directly relevant to the study of moral judgment.

Neuroeconomic studies often involve using behavioral tasks developed in the field of game theory, a mathematical approach to the study of interactive decision making (von Neumann & Morgenstern, 1944). These tasks simulate the kinds of social conditions under which people make decisions, although of course these are significantly simplified so that they can be studied in a laboratory setting. One advance from the canonical experimental designs typically employed by decision-making researchers is that these game-theoretic tasks usually involve decisions made in the context of a social interaction. Importantly, by using models of bargaining and reciprocal exchange, these tasks often capture aspects of morality perhaps more realistically than in the vignettes used by traditional moral psychology.

3.1. Bargaining

One set of tasks that has been widely used in this field is the family of dictator games (DGs) and ultimatum games (UGs). In both tasks, two players are asked to divide a sum of money between them, with this money provided by the experimenter. Both players are fully aware of the respective rules of the game, and there are no subsequent rounds on which to reach agreement (i.e., these are ''one-shot'' games). One player (the Proposer) makes a proposal to the other (the Responder) as to how this money should be divided. For instance, the Proposer may decide to divide \$10 evenly, leaving each player with \$5, or unevenly, giving \$8 to himself and \$2 to the second player.

In the DG, the Proposer decides directly how much of the endowment to award to the Responder. Allocations in this game measure pure altruism, in that the Proposer (usually) sacrifices some personal gain to share the endowment with his or her partner, often giving \$1 or \$2 to his or her partner. This result deviates from the predictions of classical game theory, which predicts that the Proposer should keep all of the money for him or herself.

In the UG, the Responder also sees the offer, but now can decide to either accept or reject this proposal. If the offer is accepted, the money is divided as suggested. However, if the offer is rejected, neither player receives anything. Classical game theory predicts that the Responder should accept any nonzero offer, on the grounds that some money is preferable to no money. The Proposer, knowing this, should therefore make the smallest possible nonzero offer to the responder, for example, a single cent. However, to no great surprise, neither of these predictions is consistent to that typically observed in the UG (see Camerer, 2003, for a useful summary of the main results). Proposers tend to offer much more than the minimum to the second player, and in fact the modal offer is usually half of the total pot. When unfair offers are made (usually defined as 20% or less of the pot), responders reject about half of the time.

Although the UG is a relatively simple game, it captures many elements of everyday social interaction, and we believe games like the UG can also be useful tools in the study of moral psychology. In particular, these games could be useful in investigations of moral values. As a working definition, moral values are taken to be culturally shaped social concepts that represent principles, standards, goals, or attitudes held by individuals or groups (Moll, Zahn, de Oliveira-Souza, Krueger, & Grafman, 2005). Concepts such as justice, honor, and fairness are common examples of moral values. Although a small number of studies (Takahashi et al., 2008; Zahn et al., 2009) have begun to investigate the neural underpinnings of moral values using paradigms in which individuals read short statements either consistent or inconsistent with common moral values (e.g., John acted generously toward Sam), economic games offer an opportunity to study how values are actually employed in relatively realistic social interactions. Although the role of moral values in these decisions has not typically been the focus of study in the neuroeconomic literature, economic games allow for the possibility of better understanding this aspect of human moral psychology that is often overlooked. Values such as fairness, justice, and honor have been largely neglected by contemporary researchers of moral psychology, which is surprising given that virtue-based theories of morality have a long history tracing back to Aristotle, and indeed are still popular today. These theories of morality focus not on normative rules or consequences that underlie morality, but instead on the cultivation of virtue and the development of a good character. Although a lengthy discussion of virtue ethics is impossible here, we introduce the topic because economic games offer the potential to study values that are deeply tied to our everyday moral life in a way that has largely gone ignored. The UG could be employed to examine distributive justice, fairness, and honor in a realistic setting.

As a first step in understanding fairness, experiments in which participants play both the UG and the DG have begun to allow researchers to tease apart the various motivations that underlie seemingly ''irrational'' economic behavior in these games (e.g., Scheres & Sanfey, 2006). For example, the Proposer's decision to make a fair offer in the UG could have two plausible rationales. By one account, Proposers might be motivated by self-interest and thus want to ensure that their offer is accepted so they can be sure of a certain gain. Alternately, Proposers could be motivated by a sense of fairness, and offer an even split because it is the right thing to do. Although both of these motivations may be involved in UG decisions, we can gain a window on behavior by comparing play in the UG to that on the DG, in which there is no need for strategic motivations as the Responder has no choice but to accept the offer. By looking at the difference between an individual's offers in these two games, the role that fairness and self-interest play in decision behavior can begin to be teased apart. Based on the data, it appears that both motivations play a role—DG offers are typically less generous than UG offers, but they rarely fall to zero. Additionally, by coupling this game-pairing strategy with individual difference measures, such as sensitivity to reward or punishment, we can also begin to uncover why moral disagreements about what is ''fair'' or ''just'' might arise so often in economic decisions.

Recent work is also beginning to shed light on how expectations affect decision behavior and potentially the flexibility of moral and social norms and values. A recent study (Sanfey, 2009) demonstrated that providing Responders with information about ''typical'' offers made in the past had a large impact on their decisions about unfair offers made to them. When Responders were informed that in general offers are quite fair, they were much more likely to reject unfair offers than if they had been told that typical offers are unfair. This demonstrates that concepts of fairness and equality are relatively easily manipulated in experimental participants, and as such suggests that these notions may be more labile than theories of moral psychology might predict.

Finally, by taking advantage of the careful mathematical modeling that is a hallmark of the economic toolbox, neuroeconomists have begun to model the complex social and emotional processes that might underlie much of moral behavior, including inequity aversion (Fehr & Schmidt, 1999), guilt (Battigalli & Dufwenberg, 2007), anger, and shame. Mathematical modeling offers a potentially very useful approach to examining individual differences underlying moral cognition and, when coupled with neuroimaging techniques, to moral neuroscience. Improved conceptual models of moral values would be an important contribution to a thorough understanding of moral cognition and could also help estimate the degree to which moral values are static or flexible in their application. In both of these endeavors, a neuroeconomic approach is likely to be particularly useful. It is still an open question to what extent people's choices in these economic games are affected by considerations of fairness or justice, but exploring this question would provide us with important new knowledge both for the study of moral psychology and neuroscience.

3.2. Reciprocal exchange

Another specific focus of game theory is to model reciprocal exchange, in which an individual provides something of value to a social partner with the expectation that the recipient will reciprocate in the future. Although greed and fear of exploitation threaten the stability of reciprocal exchange, society as a whole is more productive when reciprocity is thriving (Axelrod, 1984). Typically, reciprocal exchange is studied via the trust game (TG) and closely related prisoner's dilemma (PD) game. In the TG (Berg, Dickhaut, & McCabe, 1995), a player (the Investor) must decide how much of an endowment to invest with a partner (the Trustee). Once transferred, this money is multiplied by some factor (usually tripled or quadrupled), with the Trustee then having the opportunity to return all, some, or none of the amount back to the Investor. If the Trustee honors trust, and returns money to the Investor, both players can end up with a higher monetary payoff than was originally obtained. However, if the Trustee abuses trust and keeps the entire amount, the Investor ends up with a loss. The well-studied PD game is similar to the TG, except that both players simultaneously choose whether or not to trust each other without knowledge of their partner's choice.

These games raise additional moral issues related to cooperation and reciprocity, such as how we should treat someone who either has or has not indicated that he or she wants to place trust in us. They have also been used to examine how prior opinions of the moral status of another change when we learn additional information about those individuals. For example, in a study by Delgado, Frank, and Phelps (2005), participants saw general personality information about partners prior to playing a TG. This information consisted of vignettes regarding the moral character of the partner, with each described as either a morally positive or a morally negative person. This prior knowledge led to biases in participants' trust behavior, with a consequent reduced activity in reward-related brain areas in response to partners' game behavior. The authors' interpretation of these data is that responses to the direct actions of another can be reduced when we have been led to expect a certain pattern of behavior. This suggests that prior moral knowledge about a particular partner can reduce the degree to which we directly learn from actual behavior, and it demonstrates what has been called a ''top-down'' influence on social decision making.

These games have also been usefully employed to examine altruistic punishment, where punishment is meted out to a partner despite this accruing a cost to the punisher. Many experiments have demonstrated a widespread aversion to ''free riders,'' as evidenced by their willingness to punish them at a personal cost (Fehr & Gachter, 2002) and, more recently, by showing activation in brain reward areas when people successfully punish free riders, even at a cost to themselves, or observe them receiving punishment (De Quervain et al., 2004). This neuroeconomic approach to studying punishment allows researchers to investigate the topic in a more realistic and natural setting than the hypothetical vignettes that are typically the tool of the moral psychologists.

3.3. Emotion

In addition to the usefulness of the games themselves, recent neuroeconomic studies have begun to reveal the complex mechanisms underlying the decisions people make in these tasks, and this knowledge can be usefully applied to the study of moral judgment. In particular, these studies have also begun to reveal the important role that emotion plays in these types of decision making. Sanfey, Rilling, Aronson, Nystrom, and Cohen (2003) had participants play as the second player in the UG while undergoing fMRI and found that activity in the anterior insula, an area involved in responses to painful and disgusting stimuli, correlated strongly with the unfairness of the offer. This activation, at a group level, could predict whether responders would accept or reject the unfair offers. Additionally, insula activity was greater in games played with human, as opposed to computer, partners. In a separate study, Van't Wout, Kahn, Sanfey, and Aleman (2006) found that skin conductance, measured as a proxy for affective state, was significantly increased for unfair UG offers when compared with fair ones. As in the previous study, increased skin conductance was also predictive of offer rejections. More support comes from recent findings that individuals shown short film clips to induce sadness are more likely to reject unfair offers than controls (Harle & Sanfey, 2007). Neuroeconomics has also begun to branch out into more complex modeling of the psychological and neural mechanisms that may underlie complex moral emotions such as guilt, shame, and anger. This approach is beginning to move beyond the traditional ''reason'' versus ''emotion'' distinction that is commonly drawn by both moral psychologists and neuroeconomists, and instead focusing on how specific emotions influence decision behavior. Merging these modeling approaches with research done on moral emotions (Haidt, 2001; Moll et al., 2007; Zahn et al., 2009) offers a promising route with the potential to be beneficial for both fields.

Although this overview only scratches the surface of the varied research being carried out in the field of neuroeconomics, it provides a sense of the overall goals and methods employed in the field. It offers the opportunity to use modern neuroscientific techniques to better understand decision making at a variety of levels and to test the descriptive validity of economic models to a degree previously unavailable.

4. Integrating moral cognition and neuroeconomics

At the heart of most of the dilemmas that have typically been studied by contemporary moral psychologists is a tension between competing values: the protection of individual rights versus the maximization of the overall welfare. Cases such as the trolley problem illustrate this tension nicely, and for this reason they have been very useful in helping to uncover the basis of moral cognition. But we feel that the field has placed an unnecessary emphasis on these types of cases. Certainly the role that morality plays in everyday life is not dominated by decisions about whether to push large men in front of speeding trolleys or whether to smother our own children lest we be killed by brutal soldiers. For most people, most of the time, morality plays a much more subtle, but certainly no less important, role. Moral intuitions, emotions, and values are constant factors in the way we make normal everyday decisions. It is here, in these more subtle aspects of morality, that we think use of some neuroeconomic principles can contribute to our understanding of the structure of the moral mind. Moreover, we think it also has the potential to add to our understanding of how economic decisions are made.

As we have already indicated, our primary goal is to suggest that moral psychology and neuroeconomics have much to offer each other. By incorporating the strengths of each field into future research, the weaknesses inherent to each field can be minimized. One limitation of much of the work on moral judgment reviewed here is the reliance on hypothetical dilemmas involving highly unusual fictional situations, which often strain credulity on the part of the participant. Although these cases have been, and continue to be, very useful in helping to uncover the cognitive, emotional, and neural mechanisms underlying moral judgment, they bring with them a host of well-known problems, including low ecological validity, reliance on self-report measures, and large numbers of uncontrolled variables. Although some of these problems may be addressable through careful experiments, we worry that other aspects of moral dilemma paradigms are not resolvable.

Even worse, the dramatic and often times hard to believe cases used in many studies may constitute such a marginal aspect of our everyday moral life that they cease to tell us much about the topic at all. As an analogy, studying moral judgment through moral dilemmas might be like understanding the effects of climate on an individual by studying only individuals in Antarctica and Death Valley, or worse yet, by asking people to consider how they would act if they were to be in those climates. Although this approach would certainly tell us something, it seems far from the most productive way of conducting the research.

As we suggested previously, taking a neuroeconomic approach to the study of moral cognition opens up domains of moral life that have been largely ignored in the recent literature. The games employed in neuroeconomic research offer a potentially more useful methodology, and use of these games with the explicit aim of uncovering the cognitive and neural mechanisms underlying moral values such as fairness and justice, as well as the effect of these values on everyday decision making, has the potential to advance both fields in important ways. Of course, it is important to note that a neuroeconomic approach itself has limitations and as such does not offer a panacea for the study of moral behavior. The games typically used, although surely capturing some aspect of social behavior, are still rather artificial laboratory tasks and do not directly ask about moral issues. Although important steps have been made to incorporate emotions into the decision process, there is still considerable debate as to how these affective states are represented, and indeed to what extent they can be considered distinct processes. The primary neuroscientific method utilized by neuroeconomics, fMRI, still has limitations in terms of data acquisition in both the spatial (brain regions) and temporal (timing of processes) domains, and the particulars of the experimental setting, that of having subjects lie in a narrow, noisy, tube, are also potential impediments.

Nonetheless, we do believe, for the reasons outlined earlier, that use of the neuroeconomic approach offers a promising avenue for understanding moral processes. Recently, a series of studies have begun to shed light on the neural mechanisms of charitable giving, with these experiments providing good exemplars of research that merges moral psychology and neuroeconomics. Charitable giving offers among the clearest examples of our moral and economic concerns being brought together. Because of this, work on the neural basis of charity has led the way in an integrated moral neuroeconomics perspective. Moll et al. (2006) had participants play a sophisticated task in which they decided whether to donate money to various charities. By manipulating the costs and benefits of donation, they were able to look at charitable giving under a number of different conditions. Perhaps most intriguingly, they found that activity in the striatum, an area associated with reward, was activated both by receiving a cash reward and by giving to charity. Harbaugh, Mayr, and Burghart (2007) found similar activations, but additionally found that these effects were substantially increased when donations were voluntary (as in the case of charity) as opposed to forced (as in the case of taxation). A third study has recently used a computational model of inequity aversion in conjunction with a charitable giving task to explore notions of distributive justice (Hsu, Anen, & Quartz, 2008). In this study, individuals are faced with the choice of whether to maximize the total amount of food that they can donate (but to donate it relatively inequitably) or to donate less food overall, but in a more equitable manner. For example, one could choose to donate 20 meals to 1 specific hungry child, or offer 6 meals each to 3 such children. They found that the degree of inequity was associated with bilateral activity in the putamen, that efficiency was associated with activity in the head of the caudate, and that individual differences in their modeling data predicted individual differences in insula activity.

These studies provide excellent examples of how moral judgments and decisions can be taken out of the realm of trolleys and into the moral concerns of everyday life. From a neuroeconomic perspective, they move beyond the simple two-player games that are routinely used and show an appreciation for the wide range of values that are involved in economic decision making. As an additional benefit, they show how the integration of the two disciplines may be useful in understanding phenomena with broad social implications and may eventually aid in making tough policy decisions.

5. Conclusion

Morality plays a crucial role in our everyday decision making. We believe that a more complete understanding of the moral mind will allow us to better explain the decisions that people make. This will in turn allow for more complete theoretical models of decision making. Neuroeconomics has shed significant light on the neural basis of decision-making behavior, but it has not yet incorporated the role of morality into many of its theories. Morality places strict constraints on our decisions, and it may be one of the overriding factors in explaining why people often deviate from the rational choice models posited by economists (Tetlock, Kristel, Elson, Green, & Lerner, 2000). Economists struggle to explain why people reject profit maximizing, but unfair, choices in the UG, when from a purely financial perspective, it seems clear that any money is better than no money at all.

We have tried to make the case that by incorporating a moral psychology perspective into our neuroeconomic models, we will be able to better understand why people make the choices they do. Moreover, this perspective will allow us to approach the study of moral psychology from a more realistic perspective than is typical in the field at this time. Research bridging the gap between these fields has already begun, and we think the promising results found in these early studies bode well for models of choice more generally that consider both the moral and the economic variables present in many decisions.

Note

[1] Complicating the interpretation of these findings, Koenigs and Tranel (2007) found that VMPFC patients are more likely than controls to reject unfair offers in the ultimatum game. For a useful discussion of how these two findings can be reconciled with each other, see Moll and de Oliveira-Souza (2007a, 2007b).

References

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

Axelrod, R. M. (1984). *The evolution of cooperation*. New York: Basic Books.

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

Battigalli, P., & Dufwenberg, M. (2007). Guilt in games. *American Economic Review*, *97*(2), 170–176.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *X*, 122–142.

Bjorklund, F., Haidt, J., & Murphy, S. (2000). Moral dumbfounding: When intuition finds no reason. *Lund Psychological Reports*, *2*, 1–23.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652.

Camerer, C. (2003). Behavioural studies of strategic thinking in games. *Trends in Cognitive Sciences*, *7*(5), 225–231.

Ciaramelli, E., Muccioli, M., Ladavas, E., & di Pellegrino, G. (2007). Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Social Cognitive Affect Neuroscience*, *2*, 84–92.

Cushman, F. A., Young, L., & Hauser, M. D. (2006). The role of conscious reasoning and intuitions in moral judgment: Testing three principles of harm. *Psychological Science*, *17*(12), 1082–1089.

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York: Grosset/Putnam.

De Oliveira-Souza, R., & Moll, J. (2000). The moral brain: A functional MRI study of moral judgment. *Neurology*, *54*, A104.

De Quervain, D. J., Fischbacher, U., Treger, V., Schellhammer, M., Schnyder, U., Buck, A., & Fehr, E. (2004). The neural basis of altruistic punishment. *Science*, *305*(5688), 1254–1258.

Delgado, M. R., Frank, R. H., & Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nature Neuroscience*, *8*(11), 1611–1618.

Dorris, M. C., & Glimcher, P. W. (2004). Activity in posterior parietal cortex is correlated with the subjective desirability of an action. *Neuron*, *44*, 365–378.

Fehr, E., & Gachter, S. (2002). Altruistic punishment in humans. *Nature*, *415*, 137–140.

Fehr, E., & Schmidt, K. M. (1999). *A theory of fairness, competition, and cooperation*. Cambridge, MA: MIT Press.

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.) (2002). *Heuristics and biases: The psychology of intuitive judgment*. New York: Cambridge University Press.

Glimcher, P., Camerer, C., Fehr, E., & Poldrack, R. (2008). *Neuroeconomics: Decision making and the brain*. London: Academic Press.

Glimcher, P. W., Dorris, M., & Bayer, H. (2005). Physiological utility theory and the neuroeconomics of choice. *Games and Economic Behavior*, *52*, 213–256.

Greene, J. D. (2007). Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. *Trends in Cognitive Sciences*, *11*(8), 322–323.

Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, *107*, 1144–1154.

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

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, T. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*(5537), 2105–2108.

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

Haidt, J., & Bjorklund, F. (2008). Social intuitionists answer six questions about moral psychology. In W. Sinnott-Armstrong (Ed.), *Moral psychology, Volume 2: The cognitive science of morality: Intuition and diversity* (pp. 181–218). Cambridge, MA: MIT Press.

Harbaugh, W., Mayr, U., & Burghart, D. (2007). Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science*, *316*, 1622–1625.

Harle, K., & Sanfey, A. G. (2007). Incidental sadness biases social economic decisions in the Ultimatum Game. *Emotion*, *7*, 876–881.

Hauser, M. D., Cushman, F. A., Young, L., Jin, R., & Mikhail, J. M. (2007). A dissociation between moral judgment and justification. *Mind and Language*, *22*(1), 1–21.

Hsu, M., Anen, C., & Quartz, S. (2008). The right and the good: Distributive justice and neural encoding of equity and efficiency. *Science*, *320*, 1092–1095.

Hume, D. (1739/1978). *A treatise of human nature*. (L. Selby-Bigge & P. H. Nidditch, Eds.), Oxford, England: Clarendon Press.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.) (1982). *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263–291.

Kahneman, D., & Tversky, A. (Eds.) (2000). *Choices, values and frames*. New York: Cambridge University Press and the Russell Sage Foundation.

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

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

Kohlberg, L. (1981). *Essays on moral development, Volume 1: The philosophy of moral development*. New York: Harper Row.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167–202.

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

Moll, J., & de Oliveira-Souza, R. (2007b). Response to Greene: Moral sentiments and reason: Friends or foes? *Trends in Cognitive Science*, *11*(8), 323–324.

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

Moll, J., de Oliveira-Souza, R., Eslinger, P. J., Gramati, I. E., & Mourão-Miranda, J. (2002). The neural correlates of moral sensitivity: A functional magnetic resonance imaging investigation of basic and moral emotions. *Journal of Neuroscience*, *22*(7), 2730–2736.

Moll, J., de Oliveira-Souza, R., Garrido, G. G., Bramati, I. E., Caparelli-Daquer, E. M. A., Paiva, M. L. M. F., Zahn, R., & Grafman, J. (2007). The self as a moral agent: Linking the neural bases of social agency and moral sensitivity. *Social Neuroscience*, *2*, 336–352.

Moll, J., Eslinger, P. J., & de Oliveira-Souza, R. (2001). Frontopolar and anterior temporal cortex activation in amoral judgment task: Preliminary functional MRI results in normal subjects. *Arquivos de Neuro-Psiquiatria*, *59*(3), 657–664.

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

Moll, J., Zahn, R., de Oliveira-Souza, R., Krueger, F., & Grafman, J. (2005). The neural basis of human moral cognition. *Nature Reviews. Neuroscience*, *6*, 799–809.

von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton, NJ: Princeton Univ. Press.

Petrinovich, L., O'Neill, P., & Jorgensen, M. J. (1993). An empirical study of moral intuitions: Towards an evolutionary ethics. *Ethology and Sociobiology*, *64*, 467–478.

Piaget, J. (1965/1932). *The moral judgment of the child*. New York: Free Press.

Pochon, J. B., Riis, J., Sanfey, A. G., Nystrom, L. E., & Cohen, J. D. (2008). Functional imaging of decision conflict. *Journal of Neuroscience*, *28*, 3468–3473.

Sanfey, A. G. (2009). Expectations and social decision-making: Biasing effects of prior knowledge on Ultimatum responses. *Mind & Society*, *8*, 93–107.

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

Schaich-Borg, J., Hynes, C., Van Horn, J., Grafton, S., & Sinnott-Armstrong, W. (2006). Consequences, action, and intention as factors in moral judgments: An fMRI investigation. *Journal of Cognitive Neuroscience*, *18*(5), 803–817.

Scheres, A., & Sanfey, A. G. (2006). Individual differences in decision-making: Drive and reward responsiveness affects strategic bargaining in economic games. *Behavioral and Brain Functions*, *2*, 35.

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

Smith, A. (1759/1966). *The theory of moral sentiments*. New York: Kelly.

Takahashi, H., Kato, M., Matsuura, M., Koeda, M., Yahata, N., Suhara, T., & Okubo, Y. (2008). Neural correlates of human virtue judgment. *Cerebral Cortex*, *18*(9), 1886–1891.

Tetlock, P. E., Kristel, O., Elson, B., Green, M., & Lerner, J. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology*, *78*, 853–870.

Thompson, J. J. (1985). The trolley problem. *Yale Law Journal*, *94*, 1395–1415.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131.

Valdesolo, P., & DeSteno, D. (2006). Manipulations of emotional context shape moral judgment. *Psychological Science*, *17*(476), 476–477.

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

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

Zahn, R., Moll, J., Paiva, M., Garrido, G., Krueger, F., Huey, E., & Grafman, J. (2009). The neural basis of human social values: Evidence from functional MRI. *Cerebral Cortex*, *19*, 276–283.