Moral Judgment and Moral Neuroscience: Intersections, Definitions, and Issues

Melanie Killen and Judith Smetana

University of Maryland and University of Rochester

ABSTRACT—The moral neurosciences are an emerging area of research that has the potential to improve our understanding of the biological basis of morality. Thus far, however, research on moral neuroscience and moral judgment development has proceeded independently and with little interconnection. The current article discusses how morality has been defined and assessed in several new and prominent lines of research in the moral neurosciences and in research over the past 30 years on moral judgment development. It discusses how the 2 lines of research could be better integrated and points to strengths and limitations in the current field. Finally, suggestions are provided for further research.

KEYWORDS—morality; moral judgment; moral neuroscience; moral development; empathy

The moral neurosciences are an emerging area of research that has the potential to improve our understanding of human social and moral behavior. The field reflects a multidisciplinary approach that integrates theory and research in philosophy, psychology, economics, and neuroscience. Although recent reports have forged new ground, ambiguities have also arisen about the meaning of the findings, particularly as different areas of scholarship define and measure morality differently. Scholars have theorized about morality as an aspect of human nature since antiquity (see Aristotle), primarily debating whether morality centers on reason and rationality (Kant, 1785/1959) or emotions and affect (Hume, 1739/1969). Thus, to some extent, the diversity of definitions is not surprising.

The empirical research on moral judgment development and decision making has drawn on philosophical definitions, rooted in Kantian theories of ethics (Kohlberg, 1971; Piaget, 1932; Turiel, 1983) and, more recently, using neo-Kantian criteria for moral judgment (Gewirth, 1978). In addition, Humean theories of affect, centering theories on shame, guilt, and empathy, have been reflected in research on prosocial behavior (Eisenberg & Fabes, 1998; Eisenberg, Spinard, & Sadovsky, 2006; Hastings, Zahn-Waxler, & McShane, 2006). In the past 30 years, a wealth of empirical studies have examined children’s moral development, as reviewed in a recent handbook on moral development (Killen & Smetana, 2006).

Moral neuroscience research often addresses issues related to current and “hot” points of discussion in psychological science, such as differences between moral reasoning and moral intuitionism, cognition and emotion, or implicit and explicit forms of judgment. This research has documented the areas of the brain that are activated when individuals solve moral dilemmas, experience social exclusion, make judgments about moral character, respond to a problem involving empathy, and make decisions involving trust (Blair & Cipolotti, 2000; Delgado, Frank, & Phelps, 2005; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001; Hauser, 2006; Marshall & Fox, 2006). These analyses of the biological underpinnings of morality bear on and could enrich the work of developmental scientists investigating the emergence of moral judgments.

At the same time, we believe that advances in moral developmental research are missing from the current neuroscience research and that an integrated approach could be mutually beneficial. In this essay, we focus primarily on moral judgment rather than moral emotion. Although researchers studying moral judgment development typically consider the role of moral emotions, and particularly empathy, the two constitute a substantial area of research with a unique set of

The first author would like to thank the social neuroscience reading group at the University of Maryland for helpful discussions about morality and neuroscience, and both the authors would like to thank Noah Simon Jampol for his assistance with the literature review.

Correspondence concerning this article should be addressed to Melanie Killen, Department of Human Development, 3304 Benjamin Building, University of Maryland, College Park, MD 20742-1131; e-mail: mkillen@umd.edu or Judith Smetana at smetana@psych.rochester.edu.

© 2008, Copyright the Author(s)
Journal compilation © 2008, Society for Research in Child Development
research questions and thus is beyond the scope of this essay. We identify several new lines of research in moral neuroscience, point to strengths and limitations in the current field, and provide suggestions for further research.

WHAT COUNTS AS MORALITY?
Currently, a large gulf exists between developmental and adult research on moral reasoning and between developmental research and moral neuroscience. Most of the moral neuroscience research has been conducted with adults and with a few exceptions; little of it has drawn on research in the developmental sciences or has employed psychological definitions of morality.

Developmental science and neuroscience share a fundamental challenge in the design and selection of instruments for investigation. Because children do not share the same interpretation of the world as adults and because the words children use may not reflect the same meaning as when used by adults, developmental psychologists have established methodologies to assess children’s understanding of complex constructs. In fact, developmental scientists have formulated specific criteria as benchmarks of various phenomena (e.g., theory of mind, causal reasoning, executive functioning, societal beliefs, and biological knowledge), including morality, in order to chart emergence, change, and ontogenetic evolution. Documenting ontogenetic changes puts constraints on what can be measured and how. Likewise, the use of functional magnetic resonance imaging (fMRI) to record brain scans creates methodological constraints, particularly when measuring responses to moral dilemmas. For this reason, and others, neuroscientists have primarily studied adults, even when their questions bear on developmental issues. Moreover, to avoid measurement error, neuroscientists must ensure that participants are thinking about the stimulus material during the brain scan; otherwise, the phenomenon of interest is not recorded.

Thus, a significant challenge for both researchers studying moral development and moral neuroscience is how to measure morality. Although adult psychological studies of morality often assume that researchers and participants have a shared implicit understanding of what is meant by the term moral, developmental psychologists need to define and implement an explicit set of criteria for establishing the phenomenon under study. In the case of morality, it does not work to simply ask the child for a moral response; this would fail to capture morality as it exists in the child’s world (Piaget, 1932).

MORAL DEVELOPMENTAL RESEARCH
As with other developmental areas, researchers interested in moral judgment development typically investigate the emergence of morality, its origins, and the nature of change. In studying social and moral judgments, researchers have demonstrated that even very young children do not treat all social rules alike. In fact, moral rules (prescriptions about fairness, others’ welfare, and rights) are differentiated from conventional rules (descriptions of customs, traditions, and regulations) beginning at around age 2.5 (Smetana, 1981; Smetana & Braeges, 1990). Based on philosophical theory and empirical research, the criteria for moral rules include generalizability, impersonality, obligatoryness, and independence from authority jurisdiction and punishment mandates, which differ from social-conventional rules and personal jurisdiction (Smetana, 2006; Turiel, 1998). As an illustration, when interviewed with hypothetical scenarios, pictures, or events, most children judge hitting to be wrong. Furthermore, they will evaluate it as wrong even if there is no rule about it or if the teacher did not see it (authority jurisdiction), they will judge it as wrong at home and at school (or another country; generalizability), and they will judge it as wrong even if the rules say it is permissible. Children also give reasons reflecting the intrinsic consequences of inflicting harm on another person (“It’ll hurt him, and you shouldn’t make people sad”).

Distinguishing morality from conventions is fundamental because it indicates that prescriptive norms about justice are viewed as independent from cultural and societal norms about group functioning (Killen, Margie, & Sinno, 2006). As an example, issues identified as moral by adults include fair division of resources (money, food), equity (in the workforce, at home, in schools), equal rights, and freedom of expression. Although these issues often conflict with cultural norms and traditions, adults who view these issues as moral do so because they perceive these issues to involve fairness and justice, which are independent from group norms (Gewirth, 1978).

Many current moral philosophers and moral developmental psychologists recognize the theoretical basis of the universality of morality. The surprising finding, demonstrated in more than 100 empirical studies, is that children and adults do as well. The research has shown that children, adolescents, and adults in rural and urban contexts, in different cultures and socioeconomic statuses, and from religious and secular backgrounds share a common view that rules about fairness and others’ welfare are generalizable, not a matter of consensus, not under authority jurisdiction, and necessary to uphold (Helwig, 2006; Wainryb, 2006).

MORAL NEUROSCIENCE RESEARCH
Moral neuroscience research has drawn on philosophical theories (Kantian rationality or Humean emotionality), behaviorist theories (rewards and punishment), or layperson definitions of morality (“anything controversial,” “it’s about sexual mores,” or “it’s what people can’t agree on”). Likewise, it often relies on moral philosophical dilemmas for stimulus materials. Moral philosophy provides well-elaborated theoretical criteria—but not operational criteria—for measuring morality (Kohlberg, 1984; Piaget, 1932; Turiel, 1983). Only a handful of
moral neuroscience studies have relied on empirical research in moral psychology—and still fewer have relied on moral developmental psychology—to define morality.

An example of one frequently used hypothetical scenario in moral neuroscience is illustrative. The “trolley car” dilemma, drawn from moral philosophy, involves a decision to kill one person to save five lives (Greene et al., 2001; Hauser, 2006). In one of the first moral neuroscience experiments, 9 participants responded to 60 dilemmas while undergoing brain scanning with fMRI. Two scenarios were administered: pulling a lever on a trolley car track, which will kill one person but avoid killing five persons, and pushing a stranger off a footbridge, which will kill one person but avoid killing five persons. The experiment was framed as a test, in part, of moral developmental theories, pitting personal and impersonal morality and challenging the Kohlbergian and Piagetian accounts of rational morality.

Portions of the medial frontal gyrus, posterior cingular gyrus, and angular gyrus/bilateral were more active when deciding to push a stranger off a footbridge than when deciding to pull a lever to kill someone (Greene et al., 2001), leading the researchers to infer a biological basis for differentiating between “moral-personal” and “moral-impersonal” acts, respectively. Most individuals choose to kill one person when it entails pulling a switch but refrain when it requires pushing an individual off a bridge. That the subtle distinctions of agency or degree of contact could reflect different areas of brain activation is intriguing. Yet, a close examination of this dilemma raises numerous questions.

First, the scenario describes horrific decisions entailing psychologically disturbing alternatives; both choices will cause harm to others and would most likely be viewed as unacceptable for most people, given that they are cast in the complex role of an “executioner” (Turiel, 2007). Are these two choices distinct forms of moral decisions? They may pivot less on the personal–impersonal moral distinction, as is claimed, and more on how directly or indirectly the harm is inflicted on another person. In fact, critical studies have pointed to a number of contextual (and nonmoral) features of the situation that could distinguish the two alternatives. Researchers have demonstrated, for example, that the locus of intervention (whether the act is initiated by the agent of harm, such as throwing a bomb on a person, or the potential victim, such as throwing a bomb on a person) is central (Waldmann & Dieterich, 2007). From a moral developmental point of view, the choice of alternative actions in the trolley car scenario is not what typically makes a moral decision difficult. Rather, the difficulty arises when a cost to the self is weighed against an obligation to do the right thing regarding the treatment of others. Thus, we need to know how individuals evaluate this problem and how they justify their choices. Without this information, we cannot determine if their thinking is moral.

Second, moral developmental psychologists have demonstrated persuasively that moral judgments are not binary events (yes or no answers) and that reasons, intentions, and motives should be examined (Colby & Kohlberg, 1987; Eisenberg & Fabes, 1998; Turiel, 1998). The reasons behind the acts define whether a decision is moral. This is a basic premise in most legal systems and moral developmental research as well. Nonetheless, few moral neuroscience experiments include assessments of reasons or motives for acts. To some extent, the methods employed in moral neuroscience research follow the Greene et al. (2001) experiments using brain imaging. Subsequently, Greene et al. interpreted their findings as support for the argument that reasons are epiphenomenal and that moral intuitions are central (Greene & Haidt, 2002). Yet, “reasons” were not assessed in their original experiments. Although this debate is outside the scope of this essay, other researchers have provided critiques of the “rationality-as-epiphenomenon” position (Carpendale & Sokol, 2007; Turiel, 2007; Waldmann & Dieterich, 2007).

In fact, the trolley car dilemma is similar to Kohlberg’s moral judgment dilemmas in which an individual is asked to make a complex decision (Colby & Kohlberg, 1987). In the most famous of these, the Heinz dilemma, a husband has to decide whether to steal a drug to save his dying wife’s life. This dilemma pits a number of social considerations against one another, including the issues of stealing, property rights, marital obligations, and the value of human life. A major difference, however, is that the Heinz dilemma interviews typically take about 45 min and involve detailed probing and counterprobing to assess how individuals evaluate and reason about each component of the situations. Moral developmental researchers have extended and refined Kohlberg’s approach by creating methodologies to analyze how children evaluate straightforward acts, events, and decisions versus complex situations or decisions in which different considerations (moral and nonmoral) overlap, conflict, or are coordinated (Killen et al., 2006; Turiel, 1998).

In contrast to recent work in moral development, the trolley car paradigm also lacks an analysis of what aspects of the dilemmas individuals weigh when making their decision. Most individuals would probably want more information about the context (e.g., why are the people sitting on the tracks?). If one operates on the principle “do not harm others,” then subtle gradations in harm may matter less than the fact that both decisions involve intrinsic negative consequences. Reasoning or justifications must be assessed to understand individuals’ decisions. Moral decisions for most humans (and some would argue, nonhuman primates) involve difficult choices between alternatives that result in varying degrees of personal or group gain (or loss). This information is absent in the trolley car dilemma. Although it is true that contextual details are never fully spelled out when using vignettes, the vignettes used in developmental interviews examining distinctions between morality and social convention (e.g., Smetana, 2006) typically are highly familiar and within the realm of everyday life, enabling participants to envision the actual situation. This
differs from the trolley car scenario, which is an unusual, never-in-your-lifetime dilemma.

Moral developmental research, then, that has focused on refining definitions and measures of morality, could be useful to ensure that moral neuroscience researchers are assessing moral decisions. In the case of the trolley car paradigm, the “personal-morality” construct does not map well onto any of the constructs or findings from moral developmental research. Indeed, it blurs distinctions between moral issues (decisions about welfare, rights, and resource distribution) and personal choices and preferences, which are judged to be up to the individual and outside of moral concern or societal regulation (Nucci, 1996). This makes it difficult to determine when individuals are making moral judgments and when their judgments fall outside the realm of morality.

INTEGRATIVE APPROACHES TO MORAL JUDGMENT AND MORAL NEUROSCIENCE

As mentioned, some areas of moral neuroscience have revealed findings that contribute to moral developmental research. These include trust and moral character (Delgado et al., 2005) as well as exclusion (Eisenberg, Lieberman, & Williams, 2003). In addition, some contributions to the integration of moral development and moral neuroscience come from research on psychopathy (Blair, 1995, 2001) and autism (Leslie, Mallon, & DiCorcia, 2006).

For example, Delgado et al. (2005) examined the neural circuitry that guided college students’ decisions about who to trust during a game in which prior information about the positive or negative moral character of the players was given. Despite the fact that there were no differences in the actual behavior among the players during the game, participants trusted the positive moral player, described as someone who helped others, made personal sacrifices, and prevented harm, more than the bad or “neutral” players. Activation in the caudate nucleus, which reflects processing feedback information, was evident when playing with the neutral partner but not the good and bad partners. Participants relied less on feedback information when playing with a good partner and instead relied on their information about that partner’s moral character. Although the focus of this research was consistent with definitions of morality in developmental research, Delgado et al. refrained from explicitly defining morality or connecting their findings to developmental research, which would have been helpful. Nonetheless, the findings provide new data on the pervasiveness of moral evaluations about others’ actions.

In a series of experiments about exclusion, Eisenberg et al. (2003) defined exclusion as a moral transgression stemming from the experience of victimization. She and her colleagues used a cyberball game to monitor reactions to exclusion. Using fMRI, the researchers found that the social pain experienced by a participant when playing a virtual game in which two other players excluded him or her from the game was similar to the experience of physical pain, as demonstrated by brain activation measurements. These findings provide evidence that individuals encode negative consequences from experiencing “victimization” or social exclusion. Thus, the authors suggest that psychological harm and physical harm share some fundamental properties.

Clearly, findings about brain activation regarding pain during exclusion and physical pain are new and require close scrutiny. Thus, a recent review of the results of 10 neuroimaging studies has shown that although perceptions of pain in others are mediated by some of the same brain areas as are activated in first-hand pain, there also are some differences in patterns of neural activation (Jackson, Barth, Powell, & Lochman, 2006). Moreover, developmental researchers have shown that exclusion is more complex than this; exclusion in studies of children and adolescents involves moral (it’s unfair), societal (it’s necessary to make groups work well), and psychological (it’s my decision) considerations and reasons (Killen, Lee-Kim, McGlothin, & Stangor, 2002). Indeed, in many situations, individuals believe that exclusion is necessary and legitimate (e.g., from sports teams, from colleges, and from academia) to create well-functioning groups and organizations. Drawing on the developmental psychological literature, it would be interesting to determine whether various forms of exclusion activate similar parts of the brain.

Recent research in developmental psychopathology, augmented by fMRI studies and focusing on autistic children, has demonstrated the specificity of moral judgments. Thus far, these studies have not used fMRI, but they move closer to the type of integration we are advocating. Blair (1995) and Leslie et al. (2006) have examined distinctions among judgments about moral transgressions, social-conventional transgressions, and theory of mind in autistic children. Blair found that theory of mind was distinct from autistic children’s social and moral rule understanding; high-functioning autistic children, including those who could pass theory of mind tasks and those who could not, were both equally good at differentiating moral and social-conventional transgressions. Blair attributed autistic children’s ability to make social rule distinctions to their responses to distress cues, which are more salient for moral than for social-conventional transgressions, as the former involve a victim, whereas the latter do not.

Leslie et al. (2006) tested the influence of distress by creating scenarios in which the victim displayed “unreasonable” and “unjustified” distress (referred to as the “cry baby” condition). As do normally developing children, autistic children did not view the victim’s excessive distress as a signal to judge the act as morally wrong. Thus, distress alone was not the key factor that autistic children used to distinguish moral and social-conventional situations. As Leslie et al. suggest, using fMRI studies to determine whether different areas of the brain are activated when children (both autistic and normally developing) engage in these different tasks is one way to integrate social
neuroscience research with theories, constructs, and operational definitions from developmental science.

CONCLUSIONS AND FUTURE DIRECTIONS

The developmental research on moral judgments can benefit neuroscience research by providing a theoretical grounding for the definition, identification, and measurement of the moral domain. Likewise, moral neuroscience research has the potential to refine moral developmental theory and provide new insights about moral judgment and decision making. Moral developmental psychologists and neuroscientists also should collaborate on designing scenarios to adapt to the experimental context of fMRI studies. Research could compare how individuals evaluate acts involving negative, positive, or benign intentions toward others and how this differs if the act results in harm to another person, allocates resources in a fair way, or violates another person’s rights.

Furthermore, knowing more about the role of agency, target, intentionality, context, and motivation would facilitate our understanding of the various components of decision making that are essential for moral judgments. Neuroscience research that examines reasons and justifications for moral decisions would also contribute to theories of moral judgment and moral development. Moreover, although moral neuroscience research has the potential to illuminate our understanding of developmental processes, most of this research to date has been conducted on adults. Despite the methodological obstacles involved in conducting fMRI studies with children, the potential of neuroimaging studies to shed light on developmental processes will not be realized until children and adolescents are involved in this research.

Much is to be gained from approaches that integrate moral judgment and moral neuroscience. Just as moral developmental scientists would be remiss to ignore advances in the biological basis of morality, so too is it a mistake to ignore the rich empirical findings on moral judgment, behavior, and development. We look forward to integrative approaches that will move the field forward.

REFERENCES


