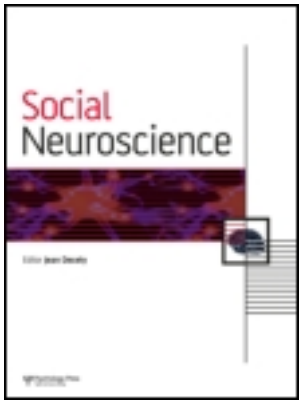

Social Neuroscience
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/psns20

Where in the brain is morality? Everywhere and maybe nowhere

Liane Young [a] & James Dungan [a]
[a] Department of Psychology, Boston College, Chestnut Hill, MA, USA

Available online: 12 Jun 2011

To cite this article: Liane Young & James Dungan (2012): Where in the brain is morality? Everywhere and maybe nowhere, Social Neuroscience, 7:1, 1-10

To link to this article: http://dx.doi.org/10.1080/17470919.2011.569146

SOCIAL NEUROSCIENCE, 2012, 7 (1), 1–10

Psychology Press
Taylor & Francis Group

Where in the brain is morality? Everywhere and maybe nowhere

Liane Young and James Dungan

Department of Psychology, Boston College, Chestnut Hill, MA, USA

The neuroscience of morality has focused on how morality works and where it is in the brain. In tackling these questions, researchers have taken both domain-specific and domain-general approaches—searching for neural substrates and systems dedicated to moral cognition versus characterizing the contributions of domain-general processes. Where in the brain is morality? On one hand, morality is made up of complex cognitive processes, deployed across many domains and housed all over the brain. On the other hand, no neural substrate or system that uniquely supports moral cognition has been found. In this review, we will discuss early assumptions of domain-specificity in moral neuroscience as well as subsequent investigations of domain-general contributions, taking emotion and social cognition (i.e., theory of mind) as case studies. Finally, we will consider possible cognitive accounts of a domain-specific morality: Does uniquely moral cognition exist?

Keywords: Moral; Emotion; Theory of mind; Domain-specificity; fMRI.

The neuroscience of morality appears to be taking an unusual course. About a decade ago, when neuroscientists first began to investigate the moral brain, the question was this: Where in the brain is morality? A key assumption was that uniquely *moral* cognition could in fact be discovered in the brain, distinct from other kinds of cognition. Researchers, using functional neuroimaging, therefore focused their efforts on constructing controlled contrasts between putatively moral versus nonmoral stimuli, attempting to uncover any selective neural response to specifically *moral* statements, scenes, and so on.

Before long, though, researchers shifted their focus from investigating "domain-specific morality" and identifying possible brain regions dedicated to morality, to characterizing the contributions of domain-general processes. Thus, rather than controlling for nonmoral dimensions like emotional and social content, moral neuroscience began to directly investigate emotion and social cognition, including reasoning about the minds of others (i.e., theory of mind) as likely inputs to moral judgment. This research revealed morality to rely on multiple domain-general processes, which are housed in many parts of the brain. Therefore, one kind of answer to the original question posed by moral neuroscientists is "everywhere"—morality is virtually everywhere in the brain. This research also represents one kind of answer to the "what" question—what is morality? In investigating the contribution of domain-general processes, moral neuroscience has begun to uncover the complex nature of morality—morality is one-part emotion, one-part theory of mind, and so on.

But the "what" question may have another kind of answer depending on how we construe the question. Morality may be many things (e.g., emotion, theory of mind), but is there any kind of cognitive content or computation that is specifically moral? Is there such a thing as a uniquely moral domain? Considering these questions requires reconsidering the possibility of domain-specificity—the possibility that, after distilling out domain-general contributions, we will

Correspondence should be addressed to: Liane Young, Department of Psychology, Boston College, McGuinn 326, Chestnut Hill, MA 02467, USA. E-mail: liane.young@bc.edu

Thanks to Jesse Prinz for his helpful comments on an earlier draft of this paper.

www.psypress.com/socialneuroscience http://dx.doi.org/10.1080/17470919.2011.569146

find uniquely moral cognition. Searching for a moral faculty in the brain, for example, will require some sense of what to look for—uniquely moral content or computation—in other words, an answer to the more basic form of the question: What is morality, and nothing else?

The structure of this review will reflect this abbreviated history of moral neuroscience. We will present early approaches in moral neuroscience and the implicit assumption, on some level, of domain-specificity. We will then discuss subsequent investigations of domain-general contributions, focusing on emotion and theory of mind as two case studies. Finally, we will return to questions concerning domain-specificity and speculate about what uniquely moral cognition might like look like in the mind and brain.

WHERE IS THE MORAL BRAIN?

Much of the early work in moral neuroscience focused on addressing the "where" question of morality—where in the brain is morality? This focus was due in part to an underlying assumption that we already had an important sense of the "what"—that is, the nature of moral content and perhaps even uniquely moral content.[1] Thus, many moral neuroscientists embarked on an investigation to locate the proprietary neural circuits of morality—neural substrates and systems dedicated to morality and not associated with any other cognitive functions, at least not yet. In aiming to identify the "moral brain," researchers appeared to be after the "where" question: Where in the brain does uniquely moral cognition reside?

Researchers attempted to address the "where" question by probing parts of the brain that could be recruited for uniquely moral cognition as opposed to any other kind of cognition. With this goal in mind, many early functional magnetic resonance imaging (fMRI) studies relied on paradigms contrasting putatively moral to nonmoral stimuli. For example, subjects read statements with moral content (e.g., "We break the law when necessary") versus nonmoral content (e.g., "Stones are made of water") (Moll, Eslinger, & de Oliveira-Souza, 2001), statements about morally good or bad actions (e.g., "A admires/steals a car") versus grammatical or ungrammatical statements (e.g., "A takes/waits a walk") (Heekeren, Wartenburger, Schmidt, Schwintowski, & Villringer, 2003), and statements describing moral violations (e.g., "He shot the victim to death") versus social violations (e.g., "He licked the dirty toilet") (Moll, de Oliveira-Souza, Eslinger et al., 2002). Activations within the ventromedial prefrontal cortex (VMPC), including the left medial orbitofrontal cortex and medial Brodmann area (BA) 10, were observed for moral versus nonmoral statements.

Other studies featured visual stimuli such as moral scenes (e.g., physical assaults, war scenes) versus nonmoral scenes (e.g., body lesions, dangerous animals), matched along a number of dimensions including emotional salience (Moll, de Oliveira-Souza, Bramati, & Grafman, 2002) and social content (i.e., number of agents or interactions depicted) (Harenski & Hamaan, 2006). Moral versus nonmoral scenes also elicited greater activation in regions of the VMPC, in particular, the right medial orbitofrontal cortex and medial frontal gyrus (BA 10 and 11) and lower medial BA 10.

Aiming to uncover neural substrates specific to the moral domain, that is, the uniquely "moral brain," this research reflected concentrated efforts to control for any confounds such as emotional salience and social content. In other words, if brain region X were recruited more for moral than nonmoral stimuli, brain region X could be implicated in specifically moral cognition only if the two stimuli sets differed along the dimension of morality and no other dimension (Poldrack, 2006). However, in spite of attempts to anticipate confounding differences, what largely emerged for moral versus nonmoral contrasts was neural evidence of greater emotional engagement and social processing. In general, moral stimuli across studies elicited greater activity in brain regions for emotional processing and social cognition, including theory of mind. These regions included not only the VMPC,[2] but also the amygdala, superior temporal sulcus (STS), bilateral temporoparietal junction (TPJ), posterior cingulate cortex (PC), and precuneus (Greene & Haidt, 2002). What this body of research uncovered, then, was not necessarily the "moral brain" but the engagement of the "emotional brain" and the "social brain" during moral cognition. As we discuss in the next sections, moral cognition is critically

[1] Throughout this review, we take moral neuroscience and psychology to address only descriptive and not prescriptive questions. The empirical project is to characterize the nature of the moral domain, the proper description of moral content and computation, not normative moral truths. Thus, we make no normative evaluation of patient populations, with cognitive deficits, whose moral judgments differ from judgments delivered by neurotypical individuals.

[2] We take fMRI activations within VMPC, including the medial portions of the orbitofrontal cortex (BA 11 and 12) as well as the medial prefrontal cortex from the ventral surface to around the level of the genu of the corpus callosum (BA 25 and portions of BA 10 and 32), to suggest the engagement of emotional processing.

supported by brain regions implicated in both emotional processing and theory of mind.

This important early work has shaped subsequent research in moral neuroscience. Students of moral cognition started to focus less on identifying a set of cognitive processes or neural substrates that might be dedicated to moral cognition and more on the very cognitive components previously identified as mere stimulus confounds—emotion and social cognition.

MORALITY IN THE EMOTIONAL BRAIN

Alongside abundant behavioral research (Greene, Morelli, Lowenberg, Nystrom, & Cohen, 2008; Inbar, Pizarro, Knobe, & Bloom, 2009; Valdesolo & DeSteno, 2006; Wheatley & Haidt, 2005), neuroscience has revealed emotion as a key input in moral judgment. Greene and colleagues were the first to investigate whether brain regions such as the VMPC,[3] implicated in emotional processing, are systematically engaged during moral judgment. Rather than contrasting moral to nonmoral stimuli, Greene and colleagues compared different kinds of moral dilemmas to each other (Greene, Sommerville, Nystrom, Darley, & Cohen, 2001). Moral dilemmas could be either "personal," that is, highly emotionally salient (e.g., pushing a man off a bridge so that his body stops a trolley from hitting five other people) or "impersonal" and less emotionally salient (e.g., turning a trolley away from five people and onto one person instead). Enhanced activation in the VMPC (medial BA 10) was associated not only with personal moral scenarios but also with personal moral judgments—judgments driven by emotional responses to harmful actions (Greene, Nystrom, Engell, Darley, & Cohen, 2004). For example, the VMPC was recruited especially for condemning personal harms like pushing the man (Greene & Haidt, 2002). Thus, brain regions like the VMPC that support emotional processing may drive one response—that is, "don't harm others"—while brain regions that support abstract reasoning and the conscious application of explicit principles (e.g., dorsolateral prefrontal cortex) may drive a different response—that is, "save as many people as possible" (Paxton & Greene, 2010). This division of labor suggests that at least part of the moral brain resides in the emotional brain.

Consistent with the neuroimaging work of Greene and colleagues, neuropsychological research has revealed abnormal moral judgment in patient populations with deficits in emotional processing. Frontotemporal dementia (FTD), for example, results in the deterioration of prefrontal and anterior temporal brain areas; FTD patients exhibit blunted emotion and diminished empathy early in the disease course (Mendez, Chen, Shapira, & Miller, 2005). Mendez and colleagues found that FTD patients were more likely to endorse personal harms like pushing the man (Mendez, Anderson, & Shapira, 2005). Like FTD patients, patients with focal VMPC lesions also exhibit blunted affect and diminished empathy, but, critically, VMPC lesion patients also retain broader intellectual function (S. W. Anderson, Barrash, Bechara, & Tranel, 2006; Barrash, Tranel, & Anderson, 2000). Therefore, any deficit reported in VMPC patients' moral judgment is due to deficits in emotional processing, and not in general cognitive functions. In two independent studies, VMPC lesion patients were more likely to endorse emotionally salient personal harms like pushing the man off the bridge to save five people—where the harm was intended and actually occurred (Ciaramelli, Muccioli, Ladavas, & di Pellegrino, 2007; Koenigs et al., 2007). Together, these results provide convergent evidence for the causally necessary role of emotional processing, subserved by the VMPC, in normal moral judgment (Prinz, 2004; Young & Koenigs, 2007).

In a follow-up study, VMPC patients showed a specific deficit in delivering moral judgments of attempted harms, including failed murder attempts—harmful intentions in the absence of harmful outcomes (Young, Bechara, et al., 2010). Specifically, VMPC patients judged attempted harms as more morally permissible compared to neurotypical control participants. VMPC patients even judged attempted harms more leniently than accidental harms—a reversal of the normal pattern of moral judgments (Cushman, 2008). Consistent with this pattern, recent fMRI work also suggests a positive correlation between VMPC activation and moral blame assigned to attempted harms (Young & Saxe, 2009b). These results suggest that VMPC patients may be unable to trigger an appropriate emotional response to information about harmful intentions, and thus they deliver moral judgments based primarily on the neutral (permissible) outcome (e.g., no harm, no foul). While the VMPC does not appear to play a role in encoding mental state information, it does appear to support emotional responses to mental state information, consistent with prior work revealing a role for the VMPC in generating emotional responses to any abstract information (Bechara & Damasio, 2005).

[3] The VMPC projects to limbic, hypothalamic, and brainstem regions that execute visceral and autonomic components of emotional responses (Ongur & Price, 2000); neurons within the VMPC encode the emotional value of sensory stimuli (Rolls, 2000).

Convergent findings from functional neuroimaging and neuropsychology suggest that emotional processing represents a crucial input to normal moral judgment, particularly in the case of agents who act on harmful intentions.[4] These findings suggest that the VMPC is a key region for emotionally mediated moral judgments.

MORALITY IN THE SOCIAL BRAIN[5]

Building on developmental psychology (Baird & Astington, 2004; Fincham & Jaspers, 1979; Piaget, 1965/1932) and recent behavioral research (Cushman, 2008; Knobe, 2005; Woolfolk, Doris, & Darley, 2006), neuroscientific work suggests that moral judgment depends on a set of social cognitive capacities that allow us to encode not only guilty intentions but also innocent intentions and, indeed, any kind of representational mental state.

Recent work has targeted the role of the "social brain" during moral cognition—how do we reason about the minds of other agents when we evaluate these agents and their actions? Guided by prior research on the neural substrates for mental state reasoning, or theory of mind, in nonmoral contexts (Jenkins & Mitchell, 2009; Perner, Aichhorn, Kronbichler, Staffen, & Ladurner, 2006; Saxe & Kanwisher, 2003; Saxe & Powell, 2006; Saxe, Schulz, & Jiang, 2006), our research suggests that a particularly selective brain region for representing mental states in the service of moral judgment is the right temporoparietal junction (RTPJ). The RTPJ appears to support distinct cognitive components of mental state reasoning for moral judgment, including the initial encoding of the agent's mental state (Young & Saxe, 2008), the use of that information for moral judgment (Young, Cushman, Hauser, & Saxe, 2007), spontaneous mental state inference (Young & Saxe, 2009a), and even post-hoc mental state reasoning to rationalize or justify moral judgments (Kliemann, Young, Scholz, & Saxe, 2008; Young, Nichols, & Saxe, 2010; Young, Scholz, & Saxe, in press). In one study, individual differences in moral judgments were correlated with individual differences in RTPJ activity (Young & Saxe, 2009b): Participants with high RTPJ activity made more lenient moral judgments of accidental harms, assigning more weight to agents' innocent intentions, versus harmful outcomes. Meanwhile, participants with a low RTPJ response made harsher moral judgments of accidents, like young children (Baird & Astington, 2004; Piaget, 1965/1932) and, as observed in a recent study, adults with high-functioning autism (Moran et al., 2011).

Disrupting RTPJ activity also disrupts the use of mental state information for moral judgment. In a recent study, we first functionally localized the RTPJ in each participant and then used transcranial magnetic stimulation (TMS) to transiently disrupt RTPJ activity during moral judgment (Young, Camprodon, Hauser, Pascual-Leone, & Saxe, 2010). Applying TMS to the RTPJ reduced the role of intentions on moral judgment and, as a direct result, increased the role of outcomes. For example, participants made more lenient moral judgments of attempted harms, relying less on information about intent, and more on information about outcome.

Notably, how exactly the RTPJ supports mental state reasoning and whether the RTPJ supports mental state reasoning specifically have been under considerable debate (Decety & Lamm, 2007). For example, some proposals suggest that the RTPJ supports low-level processes that may be engaged during mental state representation (Corbetta & Shulman, 2002; Decety & Lamm, 2007; Mitchell, 2008; Stone & Gerrans, 2006), similar to views on other candidate regions for domain-specificity (Adolphs, 2010; M. L. Anderson, in press). Indeed, regions around the RTPJ have been implicated in attention to unexpected stimuli (Corbetta, Kincade, Ollinger, McAvoy, & Shulman, 2000; Mitchell, 2008), including unexpected human actions (Buccino et al., 2007; Grezes, Frith, & Passingham, 2004; Pelphrey, Morris, & McCarthy, 2004) and inconsistent information more generally (Ferstl, Neumann, Bogler, & von Cramon, 2008; Simos, Basile, & Papanicolaou, 1997; Virtue, Parrish, & Jung-Beeman, 2008).

By contrast, we propose that the functional region of the RTPJ, reliably observed across theory-of-mind tasks (Jenkins & Mitchell, 2009; Perner et al., 2006; Saxe & Powell, 2006; Young & Saxe, 2008) for both moral judgment and standard action prediction and explanation tasks, is in fact selective for theory of mind. First, higher-resolution imaging and a bootstrap analysis revealed a small but reliable separation between the peaks of functional regions for low-level attentional processing versus theory of mind (Scholz, Triantafyllou, Whitfield-Gabrieli, Brown, &

[4] We do not take the evidence to suggest that all moral judgments are emotionally mediated. Instead, moral cognition depends on multiple inputs from multiple cognitive systems—emotional appraisals are one such input for certain kinds of moral judgments.

[5] The term "social brain" is sometimes used more generally to describe brain regions involved in any form of social cognition, including emotion. Here, we use "social brain" to refer to the brain regions involved specifically in how we reason about the minds, or mental states, of other social agents, including their beliefs and intentions (theory of mind).

Saxe, 2009). This pattern is consistent with evidence from a comprehensive meta-analysis (Decety & Lamm, 2007). Second, in a recent fMRI study, participants read stories describing unexpected or expected mental or physical states; the RTPJ response was significantly higher for mental versus physical states but not sensitive to the difference between unexpected and expected stories in the mental or the physical domain (Young, Dodell-Feder, & Saxe, 2010). Third, previous activation patterns observed for unexpected human actions have centered on the STS rather than the functional region of the RTPJ implicated in theory of mind. Notably, encountering unexpected human actions may also elicit greater theory of mind (and not just generic attention). Participants may think more about the beliefs and intentions of the actor when the actor acts unexpectedly (Buccino et al., 2007).

Together, these findings suggest that theory of mind represents a key cognitive input to moral judgment. Moral judgments depend on information about agents' beliefs and intentions. The neural substrates that support theory of mind therefore constitute an important part of the moral brain. The RTPJ is a critical node in this neural network, selectively processing mental state information during moral judgment.

WHAT IS THE MORAL BRAIN?

On the one hand, it would appear that neuroscientists have successfully answered both the "what" and the "where" of morality. Moral cognition depends on complex cognitive capacities and therefore takes up substantial space in the brain. The moral brain can be found in the emotional brain and the social brain, as we have seen, and, undoubtedly many other brain regions and brain systems that support cognitive capacities such as valuation (Shenhav & Greene, 2010), causation (Borg, Hynes, Van Horn, Grafton, & Sinnott-Armstrong, 2006; Cushman, 2008), counterfactual reasoning (Alicke, Buckingham, Zell, & Davis, 2008; Branscombe, Owen, Garstka, & Coleman, 1996; Young & Phillips, in press), agency (Gray & Wegner, 2009; Waytz, Gray, Epley, & Wegner, 2010), cognitive control (Greene et al., 2004, 2008), and deliberate, reflective reasoning (Fine, 2006; Greene et al., 2004; Kennett & Fine, 2009; Paxton & Greene, 2010). Moreover, moral judgments across different contexts, cultures, and individuals are certain to depend differently on these cognitive inputs (Cohen & Rozin, 2001; Haidt, 2001; Hamilton & Sanders, 1983; Kennett & Fine, 2009; Monin, Pizarro, & Beer, 2007).

On the other hand, emotion, theory of mind, and the cognitive capacities listed above are known to function outside moral cognition. Thus, the question remains: What is uniquely moral cognition? Is there any cognitive content or computation that is specific to morality? Is there a "moral brain" distinct from, for example, the emotional brain and the social brain? As we saw in the previous section, similar questions have been raised for theory of mind and the RTPJ (Corbetta & Shulman, 2002; Decety & Lamm, 2007; Stone & Gerrans, 2006), as well as other candidate substrates and systems for domain-specificity (M. L. Anderson, in press), including, most prominently, face processing and the fusiform face area (FFA) (Kanwisher, 2010), and social processing and the amygdala (Adolphs, 2010). By contrast, moral neuroscience has provided no candidates for substrates or systems dedicated to moral cognition. The challenge, then, is not to evaluate evidence for domain-specificity in the case of morality—there is none. Instead, we suggest that moral neuroscientists must first establish what uniquely moral cognition might look like before attempting to look for it in the brain. In the remainder of the review, we consider possible cognitive accounts of a domain-specific morality.

Early moral neuroscience, as we saw, aimed at questions along these lines, relying on a series of moral versus nonmoral contrasts while controlling for dimensions such as emotional and social content. In fact, this approach mirrors that taken by the first moral psychologists—developmental psychologists who attempted to provide a principled account of the distinctions between moral judgments and nonmoral judgments. The project was to figure out what made moral judgments *moral*. Thus, developmental psychologists studied moral cognition largely as the capacity to distinguish moral violations (e.g., hitting a classmate) from violations of social convention (e.g., wearing pajamas to class) (Turiel, 1983). Compared to conventional judgments, moral judgments were hypothesized to be *authority-independent* (i.e., an act remains immoral even when endorsed by authority), and *universal* across space and time (i.e., an immoral act is immoral everywhere, past, present, and future). For some time, the capacity to make this "moral-conventional distinction" represented *the* measure of moral capacity—not only in healthy adults but also in adults and children with developmental disorders such as psychopathy and autism (R. J. Blair, 1996; R. J. R. Blair, 1995, 1997). The moral-conventional distinction, as originally conceived, has proven to be controversial (Kelly, Stich, Haley, Eng, & Fessler, 2007; Nichols, 2002). Indeed, moral psychologists have been slow to come up with any precise characterization of what makes moral judgments specifically

moral, perhaps because this psychological boundary is especially slippery or perhaps because it just does not exist (Kelly et al., 2007; Sinnott-Armstrong, 2007).

Early fMRI studies of moral judgment assumed the existence of this boundary and proceeded to look for it in the brain. Thus, moral neuroscience and developmental psychology alike revealed an early interest in morality as its "own" domain—with possibly unique cognitive signatures and even unique neural substrates or systems. On this view, moral judgment depends on domain-general contributions but ultimately differs in detectable ways from other kinds of cognition in content or computation.

On one domain-specific account, moral judgment represents the output of a "moral faculty"—a specialized cognitive mechanism for integrating nonmoral inputs (e.g., emotions, mental state representations) in order to compute a uniquely moral judgment as its output (Hauser, 2006; Mikhail, 2002, 2007). The computation over nonmoral inputs might be simple (Cushman & Young, in press); for example, "ME HURT YOU = WRONG" (Greene et al., 2004) or "INTENT + CAUSE + HARM = WRONG" (Mikhail, 2007). And, again, cognitive representations of agency (ME), victimization (YOU), and causal and intentional attributions (CAUSE, INTENT) are widely deployed outside the moral domain (Cushman & Young, in press). On this account, what is specific to the moral domain is the "moral faculty," a mechanism that integrates the nonmoral inputs to deliver a uniquely moral judgment, which as a result, reflects the combination of the inputs (Hsu, Anen, & Quartz, 2008). Notably, there is nothing obligatory about a moral faculty—the mechanism that computes moral judgment over nonmoral inputs could operate in the same way that height and radius are taken to compute a judgment about volume (N. Anderson & Cuneo, 1978). There is no "volume" faculty distinct from "nonvolume" processes for height and radius.[6] By contrast, this account specifies the existence of an integrative mechanism that *specializes* in computing moral judgments out of nonmoral inputs.

This domain-specific account might also specify that the computation itself is uniquely moral—that is, the set of rules that are run through a moral faculty. On this account, the rules that govern moral judgment reflect a "universal moral grammar," which parallels the deep structure of language (Hauser, 2006; Mikhail, 2007; Rawls, 1971). In particular, proponents of this "linguistic analogy" argue that universal moral grammar is comprised of sets of rules that take a near-infinite array of inputs—that is, actions, causes, and intentions—to generate moral judgments.

A domain-specific account could also identify particular content within the computation as uniquely moral, rather than the computation structure, or the mechanism running the computation (e.g., moral faculty). For instance, while cognitive inputs such as CAUSE and INTENT in the cognitive rules above do not uniquely serve moral judgment (Cushman & Young, in press), it is less obvious that the content of "HURT" or "HARM" could be as widely useful outside the moral domain (S. Carey, personal communication). In fact, the content of "HURT" or "HARM" could be construed as uniquely moral and innately specified; the content could be concrete, including intentional battery (Greene et al., 2001; Mikhail, 2007) and causing personal distress (R. J. R. Blair, 1995; Leslie, Mallon, & DiCorcia, 2006), or abstract, including preventing agents from attaining their goals (Hamlin, Wynn, & Bloom, 2007; Kuhlmeier, Wynn, & Bloom, 2003). Indeed, recent developmental evidence points to these basic building blocks of morality (Bloom, 2010; Hamlin et al., 2007; Leslie et al., 2006).

Basic building blocks for morality may indeed exist, allowing even infants and toddlers to express disapproval of agents who hurt others. Somehow, these building blocks form complex moral foundations in the mature state (Haidt, 2007). If uniquely moral judgment exists, it applies not only to agents who hurt others but also agents who treat others unfairly (Bloom, 2010; van den Bos & Miedema, 2000), agents who are disloyal or disrespectful (Graham & Haidt, in press; Lowery, Unzueta, Knowles, & Goff, 2006), and agents who are impure (Borg et al., 2006; Haidt, 2007; Moll et al., 2005). Therefore, finding the uniquely moral brain would mean finding brain regions that are not only dedicated *exclusively* to moral cognition but also dedicated to *all* of moral cognition, across diverse moral contexts (e.g., harm, fairness, loyalty, respect, purity). In other words, the moral brain would have to manage *only* moral judgments and *all* moral judgments.

Before searching for a uniquely moral brain, students of moral cognition would do well to determine what, if anything, unifies morality (Sinnott-Armstrong, 2007), in the tradition of drawing the moral-conventional boundary. Can moral judgments be unified by common consequences, rather than content? For instance, while we may respond to many different kinds of moral behaviors, perhaps we always deliver some form of punishment or reward (Bastian, Jetten, & Fasoli, in press; Cushman, Dreber, Wang,

[6] Could nonmoral inputs (e.g., cause, intent) be combined in such a way that they are systematically transformed for moral judgment (Knobe, 2005)(F. Cushman, personal communication)? If so, this might provide support for a putative moral faculty.

& Costa, 2009; Henrich & Boyd, 2001). Or perhaps moral judgments always entail choices, implicit and explicit, about friendship and social alliances (Atran & Henrich, 2010; Graham & Haidt, in press; Graham, Haidt, & Nosek, 2009; Haidt, 2007; Sosis & Bressler, 2003). Taking a functional approach may ultimately inform our understanding of common cognitive components across all moral judgments.

So far, the uniquely moral brain has appeared nowhere—perhaps because it does not exist. But for those who are still interested in looking, moral psychology and neuroscience have given us a better sense of what to look for and where the challenges arise. The developmental psychologists made an early contribution to this project—the approach of characterizing unique cognitive signatures of moral judgment—what is specific to and common across all moral judgments. If the uniquely moral brain is what we are after, we might do well to follow suit. In some sense, though, we do not need to know whether something like a "moral faculty" exists to continue investigating its many cognitive inputs and their interaction. Indeed, this worthy challenge will continue to take moral psychologists all over the brain.

Original manuscript received 19 August 2010
Revised manuscript accepted 14 February 2011
First published online 16 May 2011

REFERENCES

Adolphs, R. (2010). What does the amygdala contribute to social cognition? *Annals of the New York Academy of Sciences*, *1191*, 42–61.

Alicke, M. D., Buckingham, J., Zell, E., & Davis, T. (2008). Culpable control and counterfactual reasoning in the psychology of blame. *Personality and Social Psychology Bulletin*, *34*(10), 1371–1381.

Anderson, M. L. (in press). Neural re-use as a fundamental organizational principle of the brain. *Behavioral and Brain Sciences*.

Anderson, N., & Cuneo, D. (1978). The height + width rule in children's judgments of quantity. *Journal of Experimental Psychology: General*, *107*(4), 335–378.

Anderson, S. W., Barrash, J., Bechara, A., & Tranel, D. (2006). Impairments of emotion and real-world complex behavior following childhood- or adult-onset damage to ventromedial prefrontal cortex. *Journal of the International Neuropsychological Society*, *12*(2), 224–235.

Atran, S., & Henrich, J. (2010). The evolution of religion: How cognitive by-products, adaptive learning heuristics, ritual displays, and group competition generate deep commitments to prosocial religions. *Biological Theory*, *5*(1), 18–30.

Baird, J. A., & Astington, J. W. (2004). The role of mental state understanding in the development of moral cognition and moral action. *New Directions for Child and Adolescent Development*, *103*, 37–49.

Barrash, J., Tranel, D., & Anderson, S. (2000). Acquired personality disturbances associated with bilateral damage to the ventromedial prefrontal region. *Developmental Neuropsychology*, *18*, 355–381.

Bastian, B., Jetten, J., & Fasoli, F. (in press). Cleansing the soul by hurting the flesh: The guilt-reducing effect of pain. *Psychological Science*.

Bechara, A., & Damasio, A. R. (2005). The somatic marker hypothesis: A neural theory of economic decision. *Games and Economic Behavior*, *52*(2), 336–372.

Blair, R. J. (1996). Brief report: Morality in the autistic child. *Journal of Autism and Developmental Disorders*, *26*(5), 571–579.

Blair, R. J. R. (1995). A cognitive developmental approach to morality: Investigating the psychopath. *Cognition*, *57*, 1–29.

Blair, R. J. R. (1997). Moral reasoning and the child with psychopathic tendencies. *Personality and Individual Differences*, *22*, 731–739.

Bloom, P. (2010). "The Moral Life of Babies." *New York Times Magazine*. May 9, P. MM44.

Borg, J. S., 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.

Branscombe, N., Owen, S., Garstka, T., & Coleman, J. (1996). Rape and accident counterfactuals: Who might have done otherwise and would it have changed the outcome? *Journal of Applied Social Psychology*, *26*, 1042–1067.

Buccino, G., Baumgaertner, A., Colle, L., Buechel, C., Rizzolatti, G., & Binkofski, F. (2007). The neural basis for understanding non-intended actions. *NeuroImage, 36*(Suppl 2), T119–127.

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

Cohen, A. B., & Rozin, P. (2001). Religion and the morality of mentality. *Journal of Personality and Social Psychology*, *81*(4), 697–710.

Corbetta, M., Kincade, J. M., Ollinger, J. M., McAvoy, M. P., & Shulman, G. L. (2000). Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nature Neuroscience*, *3*(3),292–297.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*(3), 201–215.

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analysis in moral judgment. *Cognition*, *108*(2), 353–380.

Cushman, F., Dreber, A., Wang, Y., & Costa, J. (2009). Accidental outcomes guide punishment in a "trembling hand" game. *PLoS One*, *4*(8), e6699.

Cushman, F., & Young, L. (in press). Patterns of moral judgment derive from non-moral psychological representations. *Cognitive Science*.

Decety, J., & Lamm, C. (2007). The role of the right temporoparietal junction in social interaction: How low-level computational processes contribute to meta-cognition. *Neuroscientist*, *13*(6), 580–593.

Ferstl, E. C., Neumann, J., Bogler, C., & von Cramon, D. Y. (2008). The extended language network: A meta-analysis of neuroimaging studies on text comprehension. *Human Brain Mapping*, *29*(5), 581–593.

Fincham, F. D., & Jaspers, J. (1979). Attribution of responsibility to the self and other in children and adults. *Journal of Personality and Social Psychology*, *37*(9), 1589–1602.

Fine, C. (2006). Is the emotional dog wagging its rational tail, or chasing it? Reason in moral judgment. *Philosophical Explorations*, *9*(1), 83–98.

Graham, J., & Haidt, J. (in press). Beyond beliefs: Religion binds individuals into moral communities. *Personality and Social Psychology Review*.

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*(5), 1029–1046.

Gray, K., & Wegner, D. M. (2009). Moral typecasting: Divergent perceptions of moral agents and moral patients. *Journal of Personality and Social Psychology*, *96*(3), 505–520.

Greene, J. D., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Science*, *6*, 517–523.

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*(3), 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*, 389–400.

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

Grezes, J., Frith, C. D., & Passingham, R. E. (2004). Inferring false beliefs from the actions of oneself and others: An fMRI study. *NeuroImage*, *21*(2), 744–750.

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. (2007). The new synthesis in moral psychology. *Science*, *316*, 998–1002.

Hamilton, V., & Sanders, J. (1983). Universals in judging wrongdoing: Japanese and Americans compared. *American Sociological Review*, *48*, 199–211.

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, *450*(7169), 557–559.

Harenski, C. L., & Hamaan, S. (2006). Neural correlates of regulating negative emotions related to moral violations. *NeuroImage*, *30*(1), 313–324.

Hauser, M. D. (2006). *Moral minds: How nature designed a universal sense of right and wrong*. New York, NY: HarperCollins.

Heekeren, H. R., Wartenburger, I., Schmidt, H., Schwintowski, H. P., & Villringer, A. (2003). An fMRI study of simple ethical decision-making. *Neuroreport*, *14*, 1215–1219.

Henrich, J., & Boyd, R. (2001). Why people punish defectors: Weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. *Journal of Theoretical Biology*, *208*, 79–89.

Hsu, M., Anen, C., & Quartz, S. R. (2008). The right and the good: Distributive justice and neural encoding of equity and efficiency. *Science*, *320*(5879), 1092–1095.

Inbar, Y., Pizarro, D. A., Knobe, J., & Bloom, P. (2009). Disgust sensitivity predicts intuitive disapproval of gays. *Emotion*, *9*(3), 435–439.

Jenkins, A. C., & Mitchell, J. P. (2009). Mentalizing under uncertainty: Dissociated neural responses to ambiguous and unambiguous mental state inferences. *Cerebral Cortex*, *20*(2), 404–410.

Kanwisher, N. (2010). Functional specificity in the human brain: A window into the functional architecture of the mind. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(25), 11163–11170.

Kelly, D., Stich, S., Haley, K., Eng, S., & Fessler, D. (2007). Harm, affect, and the moral/conventional distinction. *Mind and Language*, *22*(2), 117–131.

Kennett, J., & Fine, C. (2009). Will the real moral judgment please stand up? The implications of social intuitionist models of cognition for meta-ethics and moral psychology. *Ethical Theory and Moral Practice*, *12*(1), 77–96.

Kliemann, D., Young, L., Scholz, J., & Saxe, R. (2008). The influence of prior record on moral judgment. *Neuropsychologia*, *46*(12), 2949–2957.

Knobe, J. (2005). Theory of mind and moral cognition: Exploring the connections. *Trends in Cognitive Sciences*, *9*, 357–359.

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

Kuhlmeier, V., Wynn, K., & Bloom, P. (2003). Attribution of dispositional states by 12-month-olds. *Psychological Science*, *14*(5), 402–408.

Leslie, A. M., Mallon, R., & DiCorcia, J. A. (2006). Transgressors, victims, and cry babies: Is basic moral judgment spared in autism? *Social Neuroscience*, *1*(3–4), 270–283.

Lowery, B. S., Unzueta, M. M., Knowles, E. D., & Goff, P. A. (2006). Concern for the in-group and opposition to affirmative action. *Journal of Personality and Social Psychology*, *90*(6), 961–974.

Mendez, M., Anderson, E., & Shapira, J. (2005). An investigation of moral judgment in frontotemporal dementia. *Cognitive and Behavioral Neurology*, *18*(4), 193–197.

Mendez, M., Chen, A., Shapira, J., & Miller, B. (2005). Acquired sociopathy and frontotemporal dementia. *Dementia and Geriatric Cognitive Disorders*, *20*(2–3), 99–104.

Mikhail, J. M. (2002). *Aspects of a theory of moral cognition*. Unpublished Public Law and Legal Theory Research Paper Series. Georgetown University Law Center.

Mikhail, J. M. (2007). Universal moral grammar: Theory, evidence and the future. *Trends in Cognitive Sciences*, *11*(4), 143–152.

Mitchell, J. P. (2008). Activity in right temporo-parietal junction is not selective for theory-of-mind. *Cerebral Cortex*, *18*, 262–271.

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

Moll, J., de Oliveira-Souza, R., Eslinger, P. J., Bramati, I. E., Mourao-Miranda, J., Andreiulo, P. A., et al. (2002). The

neural correlates of moral sensitivity: A functional magnetic resonance imaging investigation of basic and moral emotions. *Journal of Neuroscience*, *22*, 2730–2736.

Moll, J., de Oliveira-Souza, R., Moll, F. T., Ignacio, F. A., Bramati, I. E., Caparelli-Daquer, E. M., et al. (2005). The moral affiliations of disgust. *Journal of Cognitive Behavioral Neurology*, *18*(1), 68–78.

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

Monin, B., Pizarro, D., & Beer, J. S. (2007). Deciding versus reacting: Conceptions of moral judgment and the reason-affect debate. *Review of General Psychology*, *11*, 99–111.

Moran, J. M., Young, L. L., Saxe, R., Lee, S. M., O'Young, D., Mavros, P. L., et al. (2011). Impaired theory of mind for moral judgment in high-functioning autism. *Proceedings of the National Academy of Sciences of the United States of America, 108*, 2688–2692.

Nichols, S. (2002). Norms with feeling: Toward a psychological account of moral judgment. *Cognition*, *84*, 221–236.

Ongur, D., & Price, J. (2000). The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys, and humans. *Cerebral Cortex*, *10*, 206–219.

Paxton, J. M., & Greene, J. (2010). Moral reasoning: Hints and allegations. *Topics in Cognitive Science*, *2*, 511–527.

Pelphrey, K. A., Morris, J. P., & McCarthy, G. (2004). Grasping the intentions of others: The perceived intentionality of an action influences activity in the superior temporal sulcus during social perception. *Journal of Cognitive Neuroscience*, *16*(10), 1706–1716.

Perner, J., Aichhorn, M., Kronbichler, M., Staffen, W., & Ladurner, G. (2006). Thinking of mental and other representations: The roles of left and right temporo-parietal junction. *Social Neuroscience*, *1*(3–4), 245–258.

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

Poldrack, R. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, *10*, 59–63.

Prinz, J. J. (2004). *Gut reactions*. New York, NY: Oxford University Press.

Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Harvard University Press.

Rolls, E. (2000). The orbitofrontal cortex and reward. *Cerebral Cortex*, *3*, 284–294.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind". *NeuroImage*, *19*(4), 1835–1842.

Saxe, R., & Powell, L. (2006). It's the thought that counts: Specific brain regions for one component of theory of mind. *Psychological Science*, *17*(8), 692–699.

Saxe, R., Schulz, L., & Jiang, L. (2006). Reading minds versus following rules: Dissociating theory of mind and executive control in the brain. *Social Neuroscience*, *1*(3–4), 284–298.

Scholz, J., Triantafyllou, C., Whitfield-Gabrieli, S., Brown, E. N., & Saxe, R. (2009). Distinct regions of right temporo-parietal junction are selective for theory of mind and exogenous attention. *PLoS One*, *4*(3), e4869.

Shenhav, A., & Greene, J. D. (2010). Moral judgments recruit domain-general valuation mechanisms to integrate representations of probability and magnitude. *Neuron*, *67*(4), 667–677.

Simos, P. G., Basile, L. F., & Papanicolaou, A. C. (1997). Source localization of the N400 response in a sentence-reading paradigm using evoked magnetic fields and magnetic resonance imaging. *Brain Research*, *762*(1–2), 29–39.

Sinnott-Armstrong, W. (2007). Is moral phenomenology unified? *Phenomenology and the Cognitive Sciences*, *7*, 85–97.

Sosis, R., & Bressler, E. (2003). Cooperation and commune longevity: A test of the costly signaling theory of religion. *Cross-Cultural Research*, *37*(2), 211–239.

Stone, V. E., & Gerrans, P. (2006). What's domain-specific about theory of mind? *Social Neuroscience*, *1*(3–4), 309–319.

Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge, UK: Cambridge University Press.

Valdesolo, P., & DeSteno, D. (2006). Manipulations of emotional context shape moral judgment. *Psychological Science*, *17*(6), 476–477.

van den Bos, K., & Miedema, J. (2000). Toward understanding why fairness matters: The influence of mortality salience on reactions to procedural fairness. *Journal of Personality and Social Psychology*, *79*(3),355–366.

Virtue, S., Parrish, T., & Jung-Beeman, M. (2008). Inferences during story comprehension: Cortical recruitment affected by predictability of events and working memory capacity. *Journal of Cognitive Neuroscience*, *20*(12), 2274–2284.

Waytz, A., Gray, K., Epley, N., & Wegner, D. M. (2010). Causes and consequences of mind perception. *Trends in Cognitive Sciences*, *14*(8), 383–388.

Wheatley, T., & Haidt, J. (2005). Hypnotic disgust makes moral judgments more severe. *Psychological Science*, *16*(10), 780–784.

Woolfolk, R. L., Doris, J. M., & Darley, J. M. (2006). Identification, situational constraint, and social cognition: Studies in the attribution of moral responsibility. *Cognition*, *100*(2), 283–301.

Young, L., Bechara, A., Tranel, D., Damasio, H., Hauser, M., & Damasio, A. (2010). Damage to ventromedial prefrontal cortex impairs judgment of harmful intent. *Neuron*, *65*, 845–851.

Young, L., Camprodon, J., Hauser, M., Pascual-Leone, A., & Saxe, R. (2010). Disruption of the right temporo-parietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgment. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 6753–6758.

Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(20), 8235–8240.

Young, L., Dodell-Feder, D., & Saxe, R. (2010). What gets the attention of the temporo-parietal junction? An fMRI investigation of attention and theory of mind. *Neuropsychologia*, *48*(9), 2658–2664.

Young, L., & Koenigs, M. (2007). Investigating emotion in moral cognition: A review of evidence from functional neuroimaging and neuropsychology. *British Medical Bulletin*, *84*, 69–79.

Young, L., Nichols, S., & Saxe, R. (2010). Investigating the neural and cognitive basis of moral luck: It's not what you do but what you know. *Review of Philosophy and Psychology*, *1*, 333–349.

Young, L., & Phillips, J. (in press). The paradox of moral focus. *Cognition*.

Young, L., & Saxe, R. (2008). The neural basis of belief encoding and integration in moral judgment. *NeuroImage*, *40*, 1912–1920.

Young, L., & Saxe, R. (2009a). An fMRI investigation of spontaneous mental state inference for moral judgment. *Journal of Cognitive Neuroscience*, *21*(7), 1396–1405.

Young, L., & Saxe, R. (2009b). Innocent intentions: A correlation between forgiveness for accidental harm and neural activity. *Neuropsychologia*, *47*(10), 2065–2072.

Young, L., Scholz, J., & Saxe, R. (in press). Neural evidence for "intuitive prosecution": The use of mental state information for negative moral verdicts. *Social Neuroscience*.