Neuron, Vol. 44, 389–400, October 14, 2004, Copyright ©2004 by Cell Press

The Neural Bases of Cognitive Conflict and Control in Moral Judgment

Joshua D. Greene,[1,2,*] Leigh E. Nystrom,[1,2] Andrew D. Engell,[1,2] John M. Darley,[1] and Jonathan D. Cohen[1,2]
[1]Department of Psychology
Princeton University
[2]Center for the Study of Brain, Mind, and Behavior
Princeton University
Princeton, New Jersey 08544

Summary

Traditional theories of moral psychology emphasize reasoning and "higher cognition," while more recent work emphasizes the role of emotion. The present fMRI data support a theory of moral judgment according to which both "cognitive" and emotional processes play crucial and sometimes mutually competitive roles. The present results indicate that brain regions associated with abstract reasoning and cognitive control (including dorsolateral prefrontal cortex and anterior cingulate cortex) are recruited to resolve difficult personal moral dilemmas in which utilitarian values require "personal" moral violations, violations that have previously been associated with increased activity in emotion-related brain regions. Several regions of frontal and parietal cortex predict intertrial differences in moral judgment behavior, exhibiting greater activity for utilitarian judgments. We speculate that the controversy surrounding utilitarian moral philosophy reflects an underlying tension between competing subsystems in the brain.

Introduction

For decades, moral psychology was dominated by developmental theories that emphasized the role of reasoning and "higher cognition" in the moral judgment of mature adults (Kohlberg, 1969). A more recent trend emphasizes the role of intuitive and emotional processes in human decision making (Damasio, 1994) and sociality (Bargh and Chartrand, 1999; Devine, 1989), a shift in perspective that has profoundly influenced recent work in moral psychology (Haidt, 2001; Rozin et al., 1999). Our previous work suggests a synthesis of these two perspectives (Greene and Haidt, 2002; Greene et al., 2001). We have argued that some moral judgments, which we call "personal," are driven largely by social-emotional responses while other moral judgments, which we call "impersonal," are driven less by social-emotional responses and more by "cognitive" processes. (As discussed below, the term "cognitive" has two distinct uses, referring in some cases to information processing in general while at other times referring to a class of processes that contrast with affective or emotional processes. Here we use quotation marks to indicate the latter usage.)

Personal moral dilemmas and judgments concern the appropriateness of personal moral violations, and we consider a moral violation to be personal if it meets three criteria: First, the violation must be likely to cause serious bodily harm. Second, this harm must befall a particular person or set of persons. Third, the harm must not result from the deflection of an existing threat onto a different party. One can think of these three criteria in terms of "ME HURT YOU." The "HURT" criterion picks out the most primitive kinds of harmful violations (e.g., assault rather than insider trading) while the "YOU" criterion ensures that the victim be vividly represented as an individual. Finally, the "ME" condition captures a notion of "agency," requiring that the action spring in a direct way from the agent's will, that it be "authored" rather than merely "edited" by the agent. Dilemmas that fail to meet these three criteria are classified as "impersonal." As noted previously (Greene et al., 2001), these three criteria reflect a provisional attempt to capture what we suppose is a natural distinction in moral psychology and will likely be revised in light of future research.

An example of an impersonal moral dilemma is the *trolley* dilemma (Thomson, 1986): A runaway trolley is headed for five people who will be killed if it proceeds on its present course. The only way to save them is to hit a switch that will turn the trolley onto an alternate set of tracks where it will kill one person instead of five. Should you turn the trolley in order to save five people at the expense of one? Most people say yes (Greene et al., 2001). An example of a personal moral dilemma is the *footbridge* dilemma (Thomson, 1986): As before, a trolley threatens to kill five people. You are standing next to a large stranger on a footbridge spanning the tracks, in-between the oncoming trolley and the hapless five. This time, the only way to save them is to push this stranger off the bridge and onto the tracks below. He will die if you do this, but his body will stop the trolley from reaching the others. Should you save the five others by pushing this stranger to his death? Most people say no (Greene et al., 2001). The trolley dilemma, unlike the footbridge dilemma, is impersonal because it involves the deflection of an existing threat (i.e., no agency—it is "editing" rather than "authoring").

The rationale for distinguishing between personal and impersonal moral violations/judgments is in part evolutionary. Evidence from observations of great apes suggests that our common ancestors lived intensely social lives guided by emotions such as empathy, anger, gratitude, jealousy, joy, love, and a sense of fairness (de Waal, 1996), and all of this in the apparent absence of moral *reasoning*. (By "reasoning" we refer to relatively slow and deliberative processes involving abstraction and at least some introspectively accessible components [Haidt, 2001].). Thus, from an evolutionary point of view, it would be strange if human behavior were not driven in part by domain-specific social-emotional dispositions. At the same time, however, humans appear

*Correspondence: jdgreene@princeton.edu

to possess a domain-general capacity for sophisticated abstract reasoning, and it would be surprising as well if this capacity played no role in human moral judgment. Thus, we sought evidence in support of the hypothesis that moral judgment in response to violations familiar to our primate ancestors (personal violations) are driven by social-emotional responses while moral judgment in response to distinctively human (impersonal) moral violation is (or can be) more "cognitive."

Our previous results supported this hypothesis in two ways (Greene and Haidt, 2002; Greene et al., 2001). First, we found that brain areas associated with emotion and social cognition (medial prefrontal cortex, posterior cingulate/precuneus, and superior temporal sulcus/temporoparietal junction) exhibited increased activity while participants considered personal moral dilemmas, while "cognitive" brain areas associated with abstract reasoning and problem solving exhibited increased activity while participants considered impersonal moral dilemmas.

Second, we found that reaction times (RTs) were, on average, considerably longer for trials in which participants judged personal moral violations to be appropriate, as compared to trials in which participants judged personal moral violations to be inappropriate. No comparable effect was observed for impersonal moral judgment. We compare this effect on RT to the Stroop effect (MacLeod, 1991; Stroop, 1935), in which people are slow to name the color of the ink in which an incongruent word appears (e.g., "red" written in green ink). According to our theory, personal moral violations elicit prepotent, negative social-emotional responses that drive people to deem such actions inappropriate. Therefore, in order to judge a personal moral violation to be appropriate one must overcome a prepotent response, just as one faced with the color-naming Stroop task must overcome the temptation to read the word "red" when it is written in green ink. The sort of mental discipline required by the Stroop task is known as "cognitive control," the ability to guide attention, thought, and action in accordance with goals or intentions, particularly in the face of competing behavioral pressures (Cohen et al., 1990; Posner and Snyder, 1975; Shiffrin and Schneider, 1977). We interpreted the behavioral results of our previous study as evidence that when participants responded in a utilitarian manner (judging personal moral violations to be acceptable when they serve a greater good) such responses not only reflected the involvement of abstract reasoning but also the engagement of cognitive control in order to overcome prepotent social-emotional responses elicited by these dilemmas.

Our present aim was to further test our theory of moral judgment by directly testing two specific hypotheses derived from the arguments above. First, we tested the hypothesis that increased RT in response to personal moral dilemmas results from the conflict associated with competition between a strong prepotent response and a response supported by abstract reasoning and the application of cognitive control. In keeping with this hypothesis, we predicted that the anterior cingulate cortex (ACC), a brain region associated with cognitive conflict in the Stroop and other tasks (Botvinick et al., 2001), would exhibit increased activity during personal moral judgment for trials in which the participant takes a long time to respond (high-RT trials), as compared to trials in which the participant responds quickly (low-RT), reflecting presumed conflict in processing. Likewise, we predicted that regions in the dorsolateral prefrontal cortex (DLPFC) would also exhibit increased activity for high-RT trials (as compared to low-RT trials), reflecting the engagement of abstract reasoning processes and cognitive control (Miller and Cohen, 2001).

Second, we tested the hypothesis that, in the dilemmas under consideration, these control processes work against the social-emotional responses described above and in favor of utilitarian judgments, i.e., judgments that maximize aggregate welfare (e.g., by sacrificing one life in order to save five others). In keeping with this hypothesis, we predicted increased DLPFC activity for trials in which participants judged personal moral violations to be appropriate, as compared to trials in which participants judged personal moral violations to be inappropriate. In other words, this hypothesis predicted that the level of activity in regions of DLPFC would correlate positively with utilitarian moral judgment. We emphasize that this prediction goes beyond those explored in our previous work. Previously, we found that different classes of moral dilemma (personal versus impersonal) produce different patterns of neural activity in the brains of moral decision makers. Here we test the hypothesis that different patterns of neural activity in response to the *same class of moral dilemma* are correlated with differences in moral decision-making behavior.

To test the predictions of this theory, we focused on a class of dilemmas that bring "cognitive" and emotional factors into more balanced tension than those featured in our previous work. For example, consider the following moral dilemma (the *crying baby* dilemma).

Enemy soldiers have taken over your village. They have orders to kill all remaining civilians. You and some of your townspeople have sought refuge in the cellar of a large house. Outside, you hear the voices of soldiers who have come to search the house for valuables.

Your baby begins to cry loudly. You cover his mouth to block the sound. If you remove your hand from his mouth, his crying will summon the attention of the soldiers who will kill you, your child, and the others hiding out in the cellar. To save yourself and the others, you must smother your child to death.

Is it appropriate for you to smother your child in order to save yourself and the other townspeople?

This is a difficult personal moral dilemma. In response to this dilemma, participants tend to answer slowly, and they exhibit no consensus in their judgments. This dilemma, like the other consistently difficult dilemmas used here, has a specific structure: in order to maximize aggregate welfare (in this case, save the most lives), one must commit a personal moral violation (in this case, smother the baby). According to our theory, this dilemma is difficult because the negative social-emotional response associated with the thought of killing one's own child competes with a more abstract, "cognitive" understanding that, in terms of lives saved/lost, one has nothing to lose (relative to the alternative) and much to gain by carrying out this horrific act. We believe that the ACC responds to this conflict and that control-related processes in the DLPFC tend to favor the aforementioned "cognitive" response. We hypothesize that these

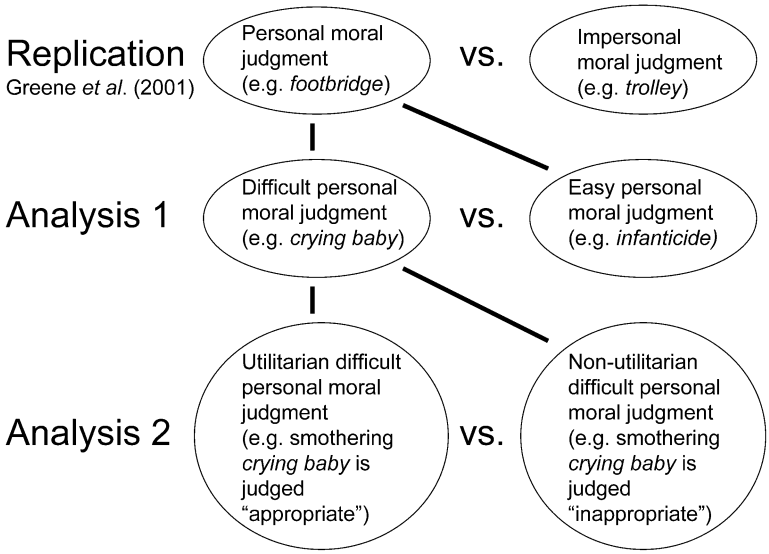

Figure 1. Relationships among Three Analyses

The present results are from three increasingly focused analyses of a single data set drawn from 41 participants who responded to moral dilemmas while having their brains scanned using fMRI.

control processes, insofar as they are effective, drive the individual to the utilitarian conclusion that it is appropriate to smother the baby in order to save more lives.

This case contrasts with "easy" personal moral dilemmas, ones that receive relatively rapid and uniform judgments (at least from the subjects within our sample). One such case is the *infanticide* dilemma in which a teenage mother must decide whether or not to kill her unwanted newborn infant. According to our theory, this dilemma is relatively easy because the negative social-emotional response associated with the thought of someone killing her own child dominates the weak or nonexistent "cognitive" case in favor of this action. Here there is no significant cognitive conflict and no need for extended reasoning or cognitive control. Thus, compared to the high-RT trials typically generated by cases like *crying baby*, the low-RT trials typically generated by cases such as *infanticide* should exhibit lower levels of activity in the ACC and DLPFC.

The analyses required to test these assertions make up a nested structure (Figure 1). Previously, we compared the neural activity associated with "personal" and "impersonal" moral judgments (Greene et al., 2001). In analysis 1, we tested our hypotheses concerning conflict monitoring in the ACC and abstract reasoning and cognitive control in the DLPFC by comparing high-RT to low-RT personal moral judgments. In analysis 2, we tested our hypothesis concerning the involvement of DLPFC in "cognitive" processes underlying utilitarian judgments by subdividing the high-RT personal moral judgments according to the participant's behavior, i.e., by comparing "utilitarian" judgments ("appropriate") to nonutilitarian judgments ("inappropriate"). In each of these difficult dilemmas, an action that normally would be judged immoral (e.g., smothering a baby) is favored by strong utilitarian considerations (e.g., saving many lives). The participants, in each instance, must decide if the utilitarian action is "appropriate" or "inappropriate." Our hypothesis is that judgments of "appropriate" will be associated with greater DLPFC activity than those of "inappropriate," reflecting the influence of "cognitive" processes favoring a utilitarian response. Analysis 2 was performed only on high-RT trials because of the relative paucity of low-RT-utilitarian judgments and the need to control for RT.

Results

Replication of Previous Results

Previously, we distinguished between personal and impersonal moral judgments and found that brain areas associated with emotion and social cognition (medial prefrontal cortex, BA 9/10; posterior cingulate/precuneus, BA 31/7; and bilateral superior temporal sulcus (STS)/inferior parietal lobe, BA 39) exhibited relatively greater activity for personal moral judgment, while brain areas associated with working memory and other characteristically "cognitive" processes (right DLPFC, BA 46; bilateral inferior parietal lobe, BA 40) exhibited relatively greater activity for impersonal moral judgment (Greene et al., 2001). The present data, drawn from 41 participants, replicated each of these results (Table 1), as well as our previously reported behavioral results. The present data set includes data from nine participants that were analyzed previously (Greene et al., 2001). A separate analysis excluding data from these nine participants yielded results consistent with those reported for the full data set. This larger data set also revealed previously unobserved differences in neural activity between personal and impersonal moral judgment (Table 1), including a bilateral increase in amygdala activity for personal, as compared to impersonal, moral judgment.

Analysis 1: Difficult versus Easy Personal Moral Judgment

Preliminary to analysis 1, we compared the neural activity associated with difficult personal moral judgments to that of a fixation baseline in ROIs generated by our previous experiment comparing personal and impersonal moral judgment (Greene et al., 2001). This was done to ensure that the difficult personal moral dilemmas focused on here (and not just the easy personal

Table 1. Brain Regions Exhibiting Differential Activity for Personal versus Impersonal Moral Judgment

Regions	Right/Left	Brodmann's Area	Max t Score (df = 40)	Cluster Size (Voxels)	Talairach Coordinates (x, y, z)
Personal > Impersonal					
Medial prefrontal cortex	R/L	9/10	10.3	513	0, 54, 19
Posterior cingulate/precuneus	R/L	31	10.01	275	−4, −50, 36
Putamen[a], caudate nucleus[a],	L	N/A, 21	7.59	320	−24, 14, −9
Middle temporal gyrus[a],					−2, −7, 8
Amygdala[a]					−1, −18, 8
					−53, −7, −13
					−26, −8, −15
Mid cingulate	R/L	24	6.17	17	0, −20, 36
Middle temporal gyrus	R	21	5.96	109	49, −4, −13
Amygdala	R	N/A	5.82	44	25, −4, −13
Anterior cingulate	R/L	24	5.47	27	−2, 22, 16
Superior temporal sulcus	R	39	5.11	233	46, −49, 19
	L	39	5.09	211	−45, −58, 17
Lingual gyrus	R	19	4.88	39	21, −65, −5
Impersonal > Personal					
Inferior parietal lobe	R	40	11.44	505	25, −64, 35
	L	40	10.02	386	−31, −61, 36
Inferior frontal gyrus	R	44	6.52	81	43, 7, 25
	L	44	6.48	77	−48, 6, 26
Posterior cingulate	R/L	23/31	6.09	20	−4, −31, 29
Middle frontal gyrus	R	46	5.97	101	39, 28, 25
Middle temporal gyrus	R	21	5.32	23	43, −49, −5
Inferior temporal gyrus	L	37	4.42	15	−40, −57, −5

Voxelwise significance threshold $p < 0.0005$; minimum cluster size 8 voxels.
[a] Denotes distinct focus of activation within larger ROI.

moral dilemmas, which previously were not distinguished from difficult personal moral dilemmas) engage the previously identified brain regions associated with emotion and social cognition. As predicted, we found that each of the three brain regions previously exhibiting greater activity for personal, as compared to impersonal, moral judgment (medial prefrontal cortex, BA 9/10; posterior cingulate/precuneus, BA 31/7; and bilateral superior temporal sulcus (STS)/inferior parietal lobe, BA 39) also exhibited above-baseline activity for difficult personal moral judgments in the current study ($p < 0.05$, cluster size ≥ 8 voxels). We note that this baseline comparison is a particularly strong test of the relevant regions' engagement in our task because these regions are most often found to exhibit decreased neural activity relative to fixation baseline in other studies (Gusnard and Raichle, 2001).

To test the hypothesis that difficult, as compared to easy, personal moral dilemmas also engage brain areas associated with abstract reasoning, cognitive conflict, and cognitive control, we directly compared the neural activity associated with difficult and easy personal moral dilemmas. More specifically, we divided personal moral judgment trials into three categories based on individually normalized reaction time (see Experimental Procedures) and compared the neural activity associated with the most difficult trials (upper third/high-RT) to the easiest trials (lower third/low-RT). Mean RT for high- and low-RT trials were 8.38 and 2.83 s, respectively. Note that the same number of time points was compared for each condition (see Experimental Procedures).

As predicted, we found that difficult, as compared to easy, personal moral dilemmas involved increased activity bilaterally in both the anterior DLPFC (BA 10/46) (see Table 2 and Figure 2) and inferior parietal lobes (BA 40/39). Also as predicted, we found that difficult, as compared to easy, personal moral dilemmas were associated with increased ACC activity (see Table 2 and Figure 2). Finally, this contrast also revealed activity in the posterior cingulate cortex (BA 23/31).

Analysis 2: Utilitarian Personal Moral Judgment

To test the hypothesis that utilitarian moral judgments engage brain areas associated with "cognitive" processes, we compared the neural activity associated with utilitarian judgments (accepting a personal moral violation in favor of a greater good) to nonutilitarian judgments (prohibiting a personal moral violation despite its utilitarian value). We conducted a planned contrast using the ROIs generated by analysis 1 (high- versus low-RT). Here we found increased activity for utilitarian, as compared to nonutilitarian, moral judgment bilaterally in the anterior DLPFC (BA 10) and in the right inferior parietal lobe (BA 40) (see Table 3 and Figure 3). In addition, we found increased activity for utilitarian moral judgments in the more anterior region of the posterior cingulate (BA 23/31) mentioned above. Other brain regions exhibit this effect as well (see Table 3). We note that the utilitarian and nonutilitarian trials compared here were matched for average RT (see Experimental Procedures).

As a supplemental exploration, we conducted a whole-brain analysis making the same comparison. As before, we observed increased activity for utilitarian judgments in the right anterior DLPFC (BA 10), inferior parietal lobe (BA 40; this time on the left side), and posterior cingulate (BA 23/31) (see Table 4 and Figure 3). This DLPFC region is anterior to and contiguous with

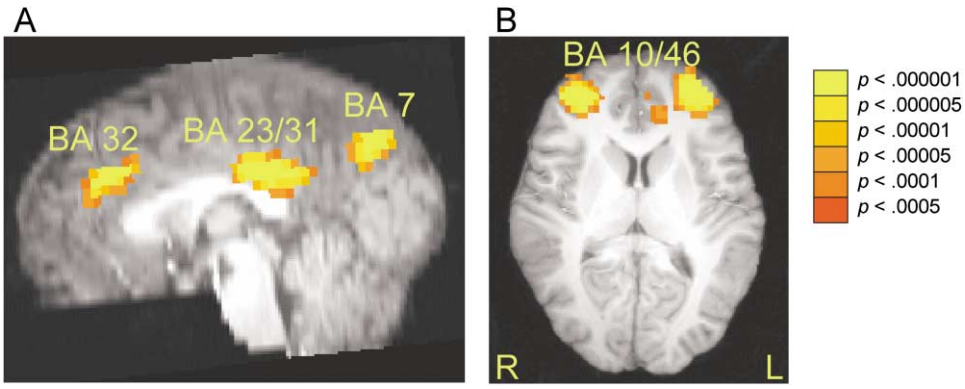

Figure 2. Difficult versus Easy Personal Moral Judgment

Selected brain regions (see Table 2) exhibiting significantly increased activity for difficult (high-RT), as compared to easy (low-RT), personal moral judgment: anterior cingulate cortex (BA 32), posterior cingulate cortex (BA 23/31), precuneus (BA 7), right and left middle frontal gyrus (BA 10/46). Statistical maps of voxelwise t scores were thresholded for significance ($p < 0.0005$) and cluster size (≥8 voxels). (A) Sagittal slice plane is $x = 0$; (B) axial slice plane is $z = +9$ (Talairach and Tournoux, 1988). Image is reversed right to left according to radiologic convention.

the DLPFC region identified in the spatially restricted analysis described above. The same effect (utilitarian > nonutilitarian) was observed in three locations within the temporal lobes: right superior temporal gyrus (BA 22/42), right middle temporal gyrus (BA 21), and left inferior temporal gyrus (BA 19).

Discussion

Analysis 1

In analysis 1, we tested two predictions concerning the psychological processes engaged in moral judgment and their neural implementation. First, according to our theory, difficult (high-RT) personal moral dilemmas such as the *crying baby* dilemma involve a conflict between (1) social-emotional responses that drive people to disapprove of personal moral violations and (2) countervailing "cognitive" processes that drive people to approve of such violations in the relevant contexts. This hypothesis stands in contrast to the hypothesis that difficult personal moral dilemmas elicit increased reaction times simply because they involve extended computation and not because of cognitive competition between incompatible behavioral responses. Evidence suggests that the ACC is responsive to processing conflict (Botvinick et al., 2001; Carter et al., 1998), as in the Stroop task (Kerns et al., 2004; MacDonald et al., 2000). We therefore predicted that difficult, as compared to easy, personal moral dilemmas would exhibit increased ACC activity, a prediction that was confirmed.

Second, we hypothesized that the processes that compete with social-emotional responses to difficult personal moral dilemmas are ones that rely on abstract reasoning and cognitive control. The anterior DLPFC is

Table 2. Brain Regions Exhibiting Differential Activity for Difficult versus Easy Personal Moral Judgment

Regions	Right/Left	Brodmann's Area	Max t Score (df = 39)	Cluster Size (Voxels)	Talairach Coordinates (x, y, z)
Difficult > Easy					
Anterior cingulate[a]	R/L	32	8.82	443	0, 33, 25
Middle frontal gyrus[a]	L	10/46			−28, 49, 7
Middle frontal gyrus[a]	R	10	7.22	213	32, 47, 11
Anterior insula[a]/inferior frontal gyrus	R	N/A, 47			32, 17, −2
Posterior cingulate[a]	R/L	23/31	7.15	380	−1, −27, 27
Precuneus[a]	R/L	7/31			0, −69, 37
Inferior parietal lobe	R	40/39	6.72	156	46, −54, 35
Inferior parietal lobe	L	40/39	5.67	40	−42, −59, 35
Anterior insula	L	N/A	4.4	9	−37, 14, 0
Easy > Difficult					
Cuneus	R/L	17/18/19	8.9	1314	1, −75, 8
Middle temporal gyrus	R	21	6.99	87	50, −6, −7
Middle temporal gyrus[a]	L	21	6.31	299	−55, −6, −7
Superior temporal sulcus[a]	L	21/22			−58, −47, 8
Superior temporal gyrus	R	22	6.22	67	46, −36, 8
Precentral gyrus	L	4	5.13	18	−52, −9, 30
Red nucleus	L	N/A	4.92	49	−4, −23, −1
Caudate nucleus	L	N/A	4.55	14	−3, 4, 8
Inferior frontal gyrus	L	44	4.15	8	−49, 5, 19

Voxelwise significance threshold $p < .0005$; min cluster size 8 voxels.
[a] Denotes distinct focus of activation within larger ROI.

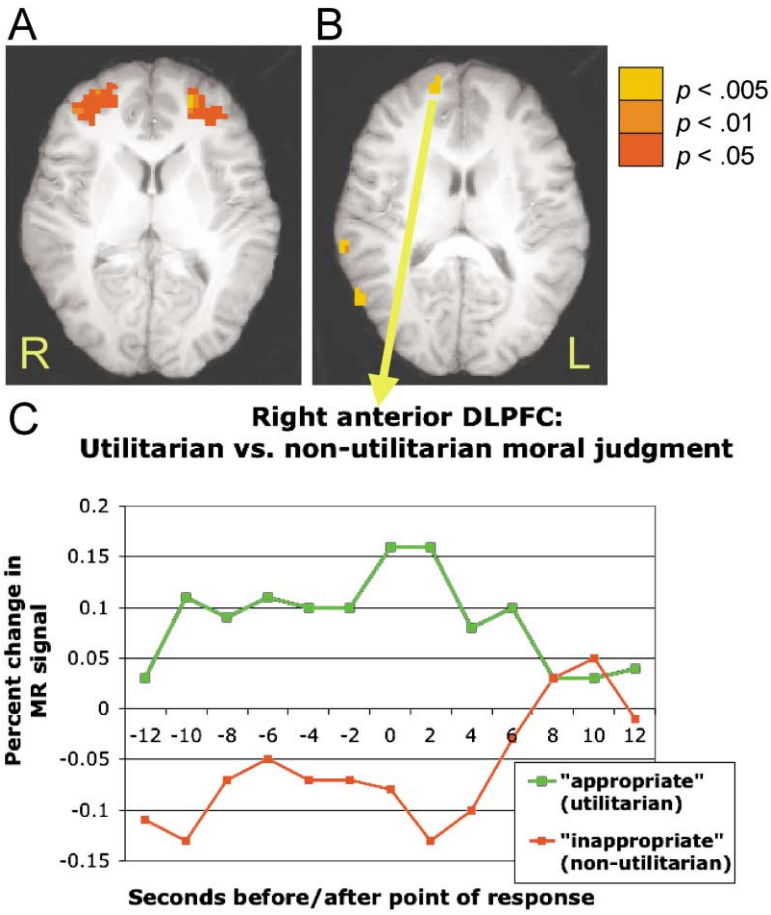

Figure 3. Utilitarian versus Nonutilitarian Difficult Personal Moral Judgment

Selected brain regions (see Tables 3–4) exhibiting significantly increased activity for utilitarian, as compared to nonutilitarian, difficult personal moral judgment. Images are reversed right to left according to radiologic convention. (A) A spatially restricted analysis ($p < 0.05$, cluster size ≥ 8) of activity in the anterior dorsolateral prefrontal cortex (BA 10/46) revealed bilateral clusters of voxels exhibiting increased activity during trials in which participants made utilitarian judgments. Axial slice plane is $z = +8$ (Talairach and Tournoux, 1988). (B) A whole-brain analysis ($p < 0.005$, cluster size ≥ 8) revealed a contiguous and slightly anterior region on the right side exhibiting the same effect ($z = +13$) . (C) Time course of activity in this region by participant response: utilitarian/"appropriate" (green) versus nonutilitarian/"inappropriate" (red). Data are not adjusted for hemodynamic lag.

known for its role in these processes (Koechlin et al., 2003; Miller and Cohen, 2001; Ramnani and Owen, 2004), and we therefore predicted that this region as well would exhibit increased activity for difficult, as compared to easy, personal moral dilemmas. This prediction was also confirmed. This result is particularly striking in light of our previous finding that personal moral judgment involved *decreased* activity in the DLPFC, as compared to impersonal moral judgment (Greene et al., 2001).

This analysis yielded several other results that deserve attention. First, our finding that difficult, as compared to easy, personal moral judgments involved increased activity bilaterally in the inferior parietal lobes is consistent with our hypothesis concerning "cognitive" processes. Activity in these regions has regularly been observed together with that of the DLPFC in tasks that engage working memory and other characteristically "cognitive" processes (Wager and Smith, 2003). Second, we observed the same effect in an anterior region of the posterior cingulate (BA 23/31). In the replication of our previous results (Table 1), this area exhibited relatively greater activation for impersonal, as compared to personal, moral judgment. Thus, the activity in this region appears once again to follow the "cognitive" pattern observed in the DLPFC and inferior parietal lobes, despite the fact that this region lies in the posterior cingu-

Table 3. Brain Regions Exhibiting Differential Activity for Utilitarian versus Nonutilitarian Personal Moral Judgment within Regions Exhibiting Differential Activity for Difficult versus Easy Personal Moral Judgment

Regions	Right/Left	Brodmann's Area	Max t Score (df = 38)	Cluster Size (Voxels)	A Priori ROI Size (Voxels)	Talairach Coordinates (x, y, z)
Utilitarian > Nonutilitarian						
Posterior cingulate	R/L	23/31	3.34	110	380[a]	0, −31, 32
Superior/middle frontal gyrus	L	10	3.32	57	443[a]	−22, 48, 8
	R	10	2.76	62	213[a]	28, 49, 6
Precuneus	L	7	3.19	11	380[a]	−14, −67, 33
	R	7	2.11	12	380[a]	7, −68, 42
Inferior parietal lobe	R	40	3.01	40	156	36, −44, 36
Lingual gyrus	R/L	18	2.57	15	1314	1, −67, 6

Voxelwise significance threshold $p < 0.05$, within a priori ROI, min cluster size 8 voxels.
[a] Includes multiple foci of activation.

Table 4. Brain Regions Exhibiting Differential Activity for Utilitarian versus Nonutilitarian Personal Moral Judgment

Regions	Right/Left	Brodmann's Area	Max t Score (df = 38)	Cluster Size (Voxels)	Talairach Coordinates (x, y, z)
Utilitarian > Nonutilitarian					
Inferior temporal gyrus	L	19	3.93	68	−51, −69, −1
Middle temporal gyrus	R	21	3.65	14	54, −57, 8
Superior frontal gyrus	R	10	3.40	8	16, 56, 12
Posterior cingulate	L/R	23/31	3.4	40	−5, −30, 32
Inferior parietal lobe	L	40	3.4	21	−38, −45, 25
Superior temporal gyrus	R	22/42	3.34	9	64, −31, 8

Voxelwise significance threshold $p < .005$; min cluster size 8 voxels.

late, a region previously associated with emotion (Maddock, 1999). (See also the discussion of analysis 2 below.) The remaining region exhibiting this effect, the anterior insula, is also associated with emotion and, more specifically, with disgust (Calder et al., 2001). Its activity has also been associated with risky decision making in a gambling task (Paulus et al., 2003) and with a tendency to reject unfair offers in the ultimatum game (Sanfey et al., 2003). Thus, it seems that the insula subserves negative affective states that bear on decision making. In light of this, it is plausible that increased insula activity would be associated with difficult personal moral judgment specifically, but not with personal moral judgment in general. In response to difficult personal moral dilemmas, people find themselves entertaining, and in some cases endorsing, actions that would otherwise be considered morally repugnant. For example, in the *crying baby* case, one is tempted to say that it is acceptable to smother the baby in order to save the lives of the other people who are hiding. The consideration of such repugnant acts in the context of difficult moral dilemmas may elicit greater insula activity than the consideration of similar acts in the context of easy personal moral dilemmas, such as the *infanticide* case, in which judgment is unanimous and swift and in which there appears to be little temptation to make a controversial judgment.

We note that dilemmas were classified as difficult (high-RT) or easy (low-RT) on a subject-by-subject, per-trial basis. Thus, a dilemma that was easy for one person could be difficult for another. Nevertheless, the RT results for each dilemma were fairly consistent across individuals, allowing us to refer to certain dilemmas as "difficult" or "easy." The *footbridge* and *infanticide* dilemmas tended to be easy, while the *crying baby* dilemma tended to be difficult. The *trolley* dilemma is impersonal and is therefore not involved in the present analysis (see Figure 1).

Several alternative explanations concerning the present analysis also deserve attention. First, the comparison between difficult and easy personal moral judgments is complicated by a potential confound of time on task: more difficult trials are defined as those associated with longer RT. However, longer RT could also reflect the prolonged engagement of other, nonspecific processes, such as visual processing and/or motor responding. We address this concern in our discussion of analysis 2 below.

Similarly, our interpretation draws on the conflict monitoring hypothesis of ACC function (Botvinick et al., 2001), rather than attributing the observed increase in ACC activity to nonspecific processes related to time on task or general difficulty. The conflict monitoring hypothesis, however, is controversial, and thus our interpretation of these data may be questioned from the perspective of alternative theories of ACC function. According to one theory, the ACC functions as an error detector (Coles et al., 1995). This theory, however, affords no clear interpretation of the present results because, in the context of difficult moral judgment, it is not clear what counts as an "error." According to a different set of theories, the relevant function of the ACC is regulative, subserving attention to (or selection for) action (e.g., Posner et al., 1988). The present data do not distinguish between such regulative theories and the conflict monitoring theory, as we have argued that difficult personal moral dilemmas involve both increased conflict (ACC) and increased regulative control (DLPFC). Elsewhere we have attempted to dissociate these processes, localizing the detection of conflict to the ACC (Botvinick et al., 1999, 2001; MacDonald et al., 2000) and control to DLPFC (MacDonald et al., 2000; Kerns et al., 2004). Finally, some have argued that increased ACC activity reflects the generation of autonomic states of cardiovascular arousal (Critchley et al., 2003), perhaps related to negative affective states. We are not able to assess this hypothesis directly because we did not acquire the necessary physiological data (heart rate variability, etc.). It is worth noting, however, that the conflict monitoring hypothesis may be consistent with the autonomic regulation hypothesis if, for example, conflict detection by the ACC is associated with negative affect and signals the need for both cognitive and autonomic control.

Analysis 2

In analysis 2, we tested a second hypothesis concerning the role of "cognitive processes" in moral judgment. In addition to proposing that difficult personal moral dilemmas involve increased reasoning and cognitive control, we hypothesized that these "cognitive" processes have a preferred behavioral outcome, namely that of favoring utilitarian moral judgments, at least in the context of the difficult personal moral dilemmas employed in this study.

These dilemmas share a common structure: a personal moral violation is required to achieve a greater good, as in the *crying baby* case. These difficult cases contrast with easy personal moral dilemmas such as the *infanticide* case in which personal moral violations

are proposed but in which the benefits sought are relatively small compared to those available in the difficult cases. Thus, it is natural to suppose that a utilitarian, cost-benefit analysis is most often the basis for judging personal moral violations to be appropriate in the difficult cases.

Reaching an overt judgment on utilitarian grounds has two processing requirements. First, the abstract reasoning that constitutes a utilitarian analysis must be conducted. Second, cognitive control must be engaged to support successful competition of the behavior favored by the outcome of that analysis against any incompatible behavioral pressures (e.g., an emotional response favoring the opposite behavior). Thus, we might expect to see neural activity associated with both of these demands in the results of analysis 1. That is, difficult personal moral dilemmas, as compared to easy ones, will involve both utilitarian reasoning and (in many cases) the application of cognitive control in favoring the utilitarian response over its competitors. However, the results of analysis 2 are expected to be more restrictive since they examined *only* difficult dilemmas, comparing activity associated with utilitarian versus nonutilitarian responses. Since difficult dilemmas were likely to have engaged abstract reasoning, irrespective of behavioral response (as evidenced by similar RTs for both types of trial), areas of activity associated with abstract reasoning are likely to have been "subtracted out" by analysis 2. Thus, voxels showing significantly greater activity associated with a utilitarian response reflect primarily the successful engagement of cognitive control in support of that response. In summary, analysis 1 identified areas within DLPFC associated with the demands for abstract reasoning as well as cognitive control, while analysis 2 identified a subset of these regions that we presume was associated more specifically with the execution of control. The latter finding is consistent with other findings suggesting that the DLPFC (bilateral BA 46) plays an important role in the regulation of potentially counterproductive emotions in the context of social decision making (Sanfey et al., 2003), in the context of the placebo effect (Wager et al., 2004), and in the evaluation of tradeoffs between future and immediate rewards (McClure et al., 2004). Our finding that different areas of PFC seem to have been differentially sensitive to the engagement of reasoning and control is intriguing, suggesting that the neural mechanisms subserving these aspects of cognitive processing may be at least partially dissociable.

The same effect was observed bilaterally in the inferior parietal lobes (BA 40), consistent with the common finding of activity in these areas in tasks engaging cognitive control (Wager and Smith, 2003). This effect was also observed in the posterior cingulate region (BA 23/31) that, as described above, exhibited increased activity for difficult personal moral dilemmas as well as increased activity for impersonal, as compared to personal, moral judgment. The activity in this region, which is more often associated with emotion (Maddock, 1999), mirrors that of the characteristically "cognitive" brain regions in the DLPFC and the inferior parietal lobes. Finally, in the whole-brain version of this analysis, this effect (utilitarian $>$ nonutilitarian) was observed in three regions within the temporal lobes. An examination of the time courses of activity suggests that the effects in these three regions were related to processes occurring after the point of decision. These may have been related to participants' reactions to their decisions, which may be more salient when the participant has recently approved of a personal moral violation (e.g., smothering a baby). In support of this suggestion, we note that overlapping regions in the temporal lobes have been associated with the perception of socially significant actions (Allison et al., 2000). The significance of the effects observed in these regions is a matter for further research.

The results of analysis 2 help to resolve a potential concern about the results of analysis 1. As noted above, the comparison made in analysis 1 between difficult and easy personal moral dilemmas, based on differences in RT, is subject to a confound of time on task. Analysis 2 used the brain areas identified in analysis 1 as a priori regions of interest in a comparison in which RT was controlled. Therefore, the results of analysis 2 suggest that the effects related to cognitive control that were predicted and observed in analysis 1 were not merely due to increased time on task and are consistent with our theory concerning the deployment of cognitive control in responding to difficult moral dilemmas.

Conflict and Cognitive Control

We hypothesized and found evidence for the engagement of brain areas associated with the detection of conflict (ACC) and the deployment of cognitive control (DLPFC) in responding to difficult moral dilemmas. In previous work, we have proposed that a cardinal function of conflict detection by ACC is the recruitment of control mechanisms in PFC needed to resolve the conflict (Botvinick et al., 2001). Support for this proposal has come from studies in which behavioral evidence of conflict and corresponding ACC activity on one trial is followed in the subsequent trial by improved performance (Botvinick et al., 1999) and a corresponding increase in PFC activity (Kerns et al., 2004). In the present study, however, difficult decisions were associated with increased ACC and DLPFC activity in the same trial. One interpretation of this finding is that the conflict associated with a difficult moral decision was detected by the ACC, which then recruited control mechanisms in DLPFC to help resolve conflict within the same trial. This is plausible given the latency of behavioral responses in this task (seconds) relative to simpler tasks involving speeded responses (under 1 s) in which adjustments of control are typically observed across trials. However, another possibility is that it was the engagement of control in the support of utilitarian responses that produced the conflict associated with difficult decisions. That is, the recruitment of control reflected in DLPFC activity allowed the utilitarian "cognitive" response to compete more effectively with the otherwise prepotent emotional response, generating the conflict reflected in ACC activity. Adjudicating between these alternatives will require greater temporal resolution than our methods provided.

"Cognition" and Emotion in Moral Psychology

For decades, moral psychology was dominated by rationalist models according to which moral development consisted of the use of increasingly sophisticated

modes of abstract moral reasoning (Kohlberg, 1969). More recently, affective processes have taken center stage. According to Haidt's social-intuitionist model (Haidt, 2001), moral judgment is driven primarily by rapid, affectively based, intuitive responses, with deliberate moral reasoning engaged after the fact to provide rational justifications in response to social demands. The present data support a synthetic theory of moral judgment according to which both of these viewpoints reflect important aspects of the truth (Greene and Haidt, 2002). Our earlier work provided initial support for this theory through our finding that personal moral judgment involves relatively greater activity in brain areas associated with social-emotional processing, while impersonal moral judgment involves relatively greater activity in brain areas associated with characteristically "cognitive" processes such as working memory, abstract reasoning, and problem solving. These neuroscientific findings were complemented by RT data suggesting that some moral dilemmas elicit response conflict between negative emotional responses and countervailing processes, which we hypothesized to be "cognitive" in nature. The present data provide further support for this theory in two ways. First, they corroborate and extend our earlier findings in a much larger sample size, providing further support for our claims concerning the roles of emotion and "cognition" in moral judgment. Second, and more importantly, these data reveal that neural activity in classically "cognitive" brain regions predicts a particular type of moral judgment behavior, thus providing strong support for the view that both "cognitive" and emotional processes play crucial and sometimes mutually competitive roles.

In analysis 1, we tested and confirmed the prediction that brain regions involved in mediating response conflict (ACC) and the implementation of cognitive control (DLPFC) exhibit increased activity during difficult, as compared to easy, personal moral judgment. These findings support the Kohlbergian claim that high-level cognitive processes are marshaled in the resolution of difficult moral dilemmas and stand in tension with the social intuitionist claim that in nearly all cases moral judgments are more akin to perception than episodes of reasoning or reflection (Haidt, 2001). Likewise, the RT data raise doubts about moral judgment as unreflective, as our participants routinely exhibited RTs over 10 s, and in some cases over 20 s, despite the fact that they were not required to justify their answers at any point. The engagement of brain areas commonly associated with deliberative thought processes strengthens this view.

In analysis 2, we tested and confirmed the prediction that utilitarian judgment, as compared to nonutilitarian judgment, involves increased activity in brain regions associated with cognitive control, particularly in the DLPFC. This finding challenges both rationalist (Kohlberg, 1969) and emotivist (Haidt, 2001) theories of moral psychology because both types of theory regard internal moral conflicts, insofar as they are thought to exist, as conflicts between processes of the same general type—rational processes in the case of rationalist models and emotional processes in the case of emotivist models. In contrast, the results presented here and previously (Greene et al., 2001) together suggest a synthetic view of moral judgment that acknowledges the crucial roles played by both emotion and "cognition" (Greene and Haidt, 2002). Our previous neuroimaging and behavioral results suggest that emotional responses drive individuals to disapprove of personal moral violations. Our present finding that increased "cognitive" activity in the DLPFC predicts utilitarian moral judgment behavior suggests that cognitive control processes can override these emotional responses, favoring personal moral violations when the benefits sufficiently outweigh the costs. Thus, both emotional and "cognitive" processes appear to be crucial in producing the patterns of neural activity and behavior observed in these experiments. This conclusion is consistent with a growing body of literature concerning the respective roles of intuition and deliberation in judgment and decision making (Kahneman, 2003).

The Relationship between "Cognition" and Emotion

The account we've offered is complicated by the fact that brain regions other than the DLPFC and inferior parietal lobes predict utilitarian moral judgment. One of these regions is in the posterior cingulate (BA 23/31), which has been associated with emotion (Maddock, 1999). This finding does not necessarily undermine our suggestion that "cognitive" processing tends to favor utilitarian judgment in response to the dilemmas employed here. It does, however, challenge the overly simple view that utilitarian judgments are wholly allied with "cognition" while nonutilitarian judgments are wholly allied with "emotion." Like David Hume (Hume, 1978), we suspect that all action, whether driven by "cognitive" judgment or not, must have some affective basis. Even a cold, calculating utilitarian must be independently motivated, first, to engage in the reasoning that utilitarian judgment requires and, second, to respond in accordance with such judgment. The ACC, a limbic region believed to recruit cognitive control (Botvinick et al., 2001), is well suited to play the first of these motivational roles. We tentatively suggest that the region identified in BA 23/31 of the posterior cingulate may play the second of these roles. This area was engaged under conditions and in a manner that closely parallels other areas (in the DLPFC and parietal cortex) that have been consistently associated with nonemotional processing. Thus, it is possible that this brain area is involved in mediating the interaction between purely "cognitive" processes and the affective/motivational processes necessary for producing behavior. This interpretation draws convergent evidence from a recent study of spatial attention (Small et al., 2003).

Throughout this article, we have relied on a familiar distinction between "emotion" or "affect" on the one hand and "cognition" on the other. This distinction has proven useful, and yet it may be somewhat artificial. The term "cognition" is often defined in terms of "information processing," but all of the processes considered here, including those that we have labeled "emotional," involve information processing, thus calling into question the usefulness of this definition of "cognition." Alternatively, one might render the emotion/cognition distinction in terms of a contrast between, on the one hand, representations that have direct motivational force and, on the other hand, representations that have no direct

motivational force of their own, but that can be contingently connected to affective/emotional states that do have such force, thus producing behavior that is both flexible and goal directed. According to this view, the emotion/cognition distinction is real, but it is a matter of degree and, at the present time, not very well understood. It is within a framework of this sort that we retain and utilize the emotion/cognition distinction while recognizing that this distinction is far from clear cut.

Broader Implications

For two centuries, Western moral philosophy has been defined largely by a tension between two opposing viewpoints. Utilitarians (or, more broadly, "consequentialists") such as John Stuart Mill (Mill, 1998) argue that morality is, or ought to be, a matter of promoting the "greater good," while "deontologists" such as Immanuel Kant (Kant, 1959) argue that certain moral lines ought not be crossed, that certain rights or duties must be respected, regardless of the greater good that might otherwise be achieved. Moral dilemmas of the sort employed here boil this philosophical tension down to its essentials and may help us understand its persistence. We propose that the tension between the utilitarian and deontological perspectives in moral philosophy reflects a more fundamental tension arising from the structure of the human brain. The social-emotional responses that we've inherited from our primate ancestors (due, presumably, to some adaptive advantage they conferred), shaped and refined by culture bound experience, undergird the absolute prohibitions that are central to deontology. In contrast, the "moral calculus" that defines utilitarianism is made possible by more recently evolved structures in the frontal lobes that support abstract thinking and high-level cognitive control. We have adduced some support for this hypothesis, albeit using a limited set of testing materials. First, we have seen evidence of increased social-emotional processing in cases in which deontological intuitions are prominent. Second, we have seen greater activity in brain regions associated with cognitive control when utilitarian judgments prevail. These brain regions house some of our species' most recently evolved neural features (Allman et al., 2002) and associated cognitive abilities.

We emphasize that this cognitive account of the Kant versus Mill problem in ethics is speculative. Should this account prove correct, however, it will have the ironic implication that the Kantian, "rationalist" approach to moral philosophy is, psychologically speaking, grounded not in principles of pure practical reason, but in a set of emotional responses that are subsequently rationalized (Haidt, 2001). Whether this psychological thesis has any normative implications is a complicated matter that we leave for treatment elsewhere (Greene, 2003; Greene and Cohen, 2004).

Experimental Procedures

Participants

Our participants were 41 healthy adult undergraduates (24 males, 17 females). All participants spoke native English, were right handed, and were screened for a history of psychiatric and neurological problems. All experimental procedures complied with the guidelines of Princeton University's Internal Review Panel. Written informed consent was obtained for each participant. In addition to the data drawn from these 41 participants, data from three participants were discarded for technical reasons related to the scanner, and data from one participant were discarded due to highly abnormal behavioral responses. One participant's data were discarded from analyses 1 and 2 and another participant's data were discarded from analysis 2 due to unbalanced factors (which depend on the participants RTs and decision outcomes).

Stimuli

We employed a battery of 60 practical dilemmas, available online at http://www.neuron.org/cgi/content/full/44/2/389/DC1/. (See materials for "Experiment 2".) These dilemmas were divided into "moral" and "nonmoral" categories based on the responses of pilot participants. Data concerning nonmoral dilemmas were not analyzed for present purposes. Two independent coders evaluated each moral dilemma using three criteria designed to capture the difference between the intuitively "up close and personal" (and putatively more emotional) sort of violation exhibited by the footbridge dilemma and the more intuitively impersonal (and putatively less emotional) violation exhibited by the trolley dilemma. First, coders indicated for each dilemma whether or not the action in question could "reasonably be expected to lead to serious bodily harm." Second, they were asked to indicate whether or not this harm would be "the result of deflecting an existing threat onto a different party." Our use of this criterion, which parallels a distinction made by Thomson (1986), is an attempt to operationalize an intuitive notion of "agency." Intuitively, when a harm is produced by means of deflecting an existing threat, the agent has merely "edited" and not "authored" the resulting harm. Finally, coders were asked to indicate whether or not the resulting harm would "befall a particular person or a member or members of a particular group of people." Here the question, in intuitive terms, is whether the victim(s) is/are "on stage" in the dilemma. The moral dilemmas of which the coders said that the action in question (1) could reasonably be expected to lead to serious bodily harm (2) to a particular person or a member or members of a particular group of people (3) where this harm is not the result of deflecting an existing threat onto a different party were assigned to the personal moral judgment condition, the others to the impersonal condition.

For analysis 1, RTs were normalized by reexpressing each RT as a percentage of the participant's average RT for personal trials. Personal trials were ordered by normalized RT and split evenly into high-, medium-, and low-RT categories. Medium-RT dilemmas were discarded from analysis. For analysis 2, a two-tailed Student's t test was performed to ensure that the utilitarian and nonutilitarian trial groups were matched for average RT ($p > 0.6$).

Experimental Design

Dilemmas were presented on a visual display projected into the scanner in a series of 12 blocks of five trials each. Each dilemma was presented as text through a series of three screens, the first two describing a scenario and the last posing a question about the appropriateness of an action one might perform in that scenario (e.g., turning the trolley). Subjects read at their own pace, pressing a button to advance from the first to the second screen and from the second to the third screen. After reading the third screen, subjects responded by pressing one of two buttons ("appropriate" or "inappropriate"). Subjects were given a maximum of 46 s to read all three screens and respond. The intertrial interval (ITI) lasted for a minimum of 14 s (seven images) in each trial, allowing the hemodynamic response to return to baseline after each trial. During the ITI, participants viewed a fixation cross. Stimuli were presented and behavioral responses were collected using PsyScope stimulus presentation software (Cohen et al., 1993), the PsyScope button box (http://psyscope.psy.cmu.edu/bbox/), and a custom response handset from Psychology Software Tools (http://www.pstnet.com/).

fMRI Acquisition

Images were acquired using a 3.0 T Siemens Allegra head-dedicated scanner. A high-resolution (1 mm voxel MPRAGE) whole-brain structural scan was acquired prior to the functional imaging. Functional images were acquired in 22 axial slices parallel to the AC-PC line

using an EPI pulse sequence, with a TR of 2000 ms, a TE of 25 ms, a flip angle of 90, a FOV of 192 mm, 3.0 mm isotropic voxels, and 1 mm interslice spacing.

fMRI Preprocessing and Analysis

Prior to statistical analysis, images for all participants were motion corrected and coregistered using AIR (Woods et al., 1992). Statistical analyses and related preprocessing were performed using the NIS software package (http://kraepelin.wpic.pitt.edu/nis/). Images were spatially smoothed with an 8 mm FWHM 3D Gaussian filter. Linear trends were removed from each voxel's time series for each run, and outliers beyond three standard deviations of the mean were corrected. Task-related activity was measured using a floating window of eight images surrounding (four prior to, one during, and three following) the time of response. Thus, the entire window was 16 s long. This window included three postresponse images in order to allow for the lag in the hemodynamic response (typically peaking 4–5 s following an eliciting neural response). For each participant, the mean BOLD signal in each condition was averaged across the eight images in the response window; these means were then contrasted across subjects with a pairwise t test.

Whole-brain exploratory analyses were performed with a voxelwise significance threshold of $p < 0.0005$, while analyses focusing on a restricted number of a priori ROI voxels were performed with a reduced voxelwise significance threshold of $p < 0.05$. All analyses used a cluster size threshold (8 voxels) to reduce type 1 error from multiple comparisons (Forman et al., 1995). The analysis preliminary to analysis 1 was restricted to voxels contained within a priori ROIs generated by experiment 1 from Greene et al. (2001). For the spatially restricted version of analysis 2, an ROI mask generated by analysis 1 (high-RT versus low-RT) was used to delimit a priori ROIs. The subsequent exploratory whole-brain analysis was performed at a lower statistical threshold ($p < 0.005$) after an initial analysis at our standard threshold ($p < 0.0005$) failed to identify any significant voxels. For generation of the event-related average graph (Figure 3C), baseline activity was defined as the mean signal across the first two images of the trial, reflecting activity during the prior ITI.

Acknowledgments

This research was supported by a fellowship from the National Institutes of Health awarded to J.D.G. (MH067410) and by grants from the National Science Foundation awarded to J.D.C, J.M.D., and J.D.G. (BCS-0351996) and to Paul Kantor and Stephen J. Hanson (CCF-0205178). Many thanks to Jason Chein and Sam McClure for helpful suggestions.

Received: June 3, 2004
Revised: September 13, 2004
Accepted: September 27, 2004
Published: October 13, 2004

References

Allison, T., Puce, A., and McCarthy, G. (2000). Social perception from visual cues: Role of the sts region. Trends Cogn. Sci. *4*, 267–278.

Allman, J., Hakeem, A., and Watson, K. (2002). Two phylogenetic specializations in the human brain. Neuroscientist *8*, 335–346.

Bargh, J.A., and Chartrand, T.L. (1999). The unberable automaticity of being. Am. Psychol. *54*, 462–479.

Botvinick, M., Nystrom, L.E., Fissell, K., Carter, C.S., and Cohen, J.D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. Nature *402*, 179–181.

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

Calder, A.J., Lawrence, A.D., and Young, A.W. (2001). Neuropsychology of fear and loathing. Nat. Rev. Neurosci. *2*, 352–363.

Carter, C.S., Braver, T.S., Barch, D.M., Botvinick, M.M., Noll, D., and Cohen, J.D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. Science *280*, 747–749.

Cohen, J.D., Dunbar, K., and McClelland, J.L. (1990). On the control of automatic processes: A parallel distributed processing account of the stroop effect. Psychol. Rev. *97*, 332–361.

Cohen, J.D., MacWhinney, B., Flatt, M., and Provost, J. (1993). Psyscope: A new graphic interactive environment for designing psychology experiments. Behavioral Research Methods, Instruments, and Computers *25*, 257–271.

Coles, M.G., Scheffers, M.K., and Fournier, L. (1995). Where did you go wrong? Errors, partial errors, and the nature of human information processing. Acta Psychol. (Amst.) *90*, 129–144.

Critchley, H.D., Mathias, C.J., Josephs, O., O'Doherty, J., Zanini, S., Dewar, B.K., Cipolotti, L., Shallice, T., and Dolan, R.J. (2003). Human cingulate cortex and autonomic control: Converging neuroimaging and clinical evidence. Brain *126*, 2139–2152.

Damasio, A.R. (1994). Descartes' Error: Emotion, Reason, and the Human Brain (New York: G.P. Putnam).

de Waal, F. (1996). Good Natured: The Origins of Right and Wrong in Humans and Other Animals (Cambridge, MA: Harvard University Press).

Devine, P.G. (1989). Stereotypes and prejudice: Their automatic and controlled components. J. Pers. Soc. Psychol. *56*, 5–18.

Forman, S.D., Cohen, J.D., Fitzgerald, M., Eddy, W.F., Mintun, M.A., and Noll, D.C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fmri): Use of a cluster-size threshold. Magn. Reson. Med. *33*, 636–647.

Greene, J. (2003). From neural 'is' to moral 'ought': What are the moral implications of neuroscientific moral psychology? Nat. Rev. Neurosci. *4*, 846–849.

Greene, J., and Cohen, J. (2004). For the law, neuroscience changes nothing and everything. Philos. Trans. R. Soc. Lond. B Biol. Sci., in press.

Greene, J., and Haidt, J. (2002). How (and where) does moral judgment work? Trends Cogn. Sci. *6*, 517–523.

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

Gusnard, D.A., and Raichle, M.E. (2001). Searching for a baseline: Functional imaging and the resting human brain. Nat. Rev. Neurosci. *2*, 685–694.

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

Hume, D. (1978). A Treatise of Human Nature, L.A. Selby-Bigge and P.H. Nidditch, eds. (Oxford: Oxford University Press).

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. Am. Psychol. *58*, 697–720.

Kant, I. (1959). Foundation of the Metaphysics of Morals (Indianapolis, IN: Bobbs-Merrill).

Kerns, J.G., Cohen, J.D., MacDonald, A.W., 3rd, Cho, R.Y., Stenger, V.A., and Carter, C.S. (2004). Anterior cingulate conflict monitoring and adjustments in control. Science *303*, 1023–1026.

Koechlin, E., Ody, C., and Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. Science *302*, 1181–1185.

Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In Handbook of Socialization Theory and Research, D.A. Goslin, ed. (Chicago: Rand McNally), pp. 347–480.

MacDonald, A.W., 3rd, Cohen, J.D., Stenger, V.A., and Carter, C.S. (2000). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. Science *288*, 1835–1838.

MacLeod, C.M. (1991). Half a century of research on the stroop effect: An integrative review. Psychol. Bull. *109*, 163–203.

Maddock, R.J. (1999). The retrosplenial cortex and emotion: New insights from functional neuroimaging of the human brain. Trends Neurosci. *22*, 310–316.

McClure, S.M., Laibson, D.I., Loewenstein, G., and Cohen, J.D. (2004). Separate neural systems value immediate and delayed monetary rewards. Science, in press.

Mill, J.S. (1998). Utilitarianism, R. Crisp, ed. (New York: Oxford University Press).

Miller, E.K., and Cohen, J.D. (2001). An integrative theory of prefrontal cortex function. Annu. Rev. Neurosci. *24*, 167–202.

Paulus, M.P., Rogalsky, C., Simmons, A., Feinstein, J.S., and Stein, M.B. (2003). Increased activation in the right insula during risk-taking decision making is related to harm avoidance and neuroticism. Neuroimage *19*, 1439–1448.

Posner, M.I., and Snyder, C.R.R. (1975). Attention and cognitive control. In Information Processing and Cognition, R.L. Solso, ed. (Hillsdale, NJ: Erlbaum), pp. 55–85.

Posner, M.I., Petersen, S.E., Fox, P.T., and Raichle, M.E. (1988). Localization of operations in the human brain. Science *240*, 1627–1631.

Ramnani, N., and Owen, A.M. (2004). Anterior prefrontal cortex: Insights into function from anatomy and neuroimaging. Nat. Rev. Neurosci. *5*, 184–194.

Rozin, P., Lowery, L., Imada, S., and Haidt, J. (1999). The cad triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). J. Pers. Soc. Psychol. *76*, 574–586.

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

Shiffrin, R.M., and Schneider, W. (1977). Controlled and automaitc information processing: Ii. Perceptual learning, automatic attending, and a general theory. Psychol. Rev. *84*, 127–190.

Small, D.M., Gitelman, D.R., Gregory, M.D., Nobre, A.C., Parrish, T.B., and Mesulam, M.M. (2003). The posterior cingulate and medial prefrontal cortex mediate the anticipatory allocation of spatial attention. Neuroimage *18*, 633–641.

Stroop, J.R. (1935). Studies of interference in serial verbal reactions. J. Exp. Psychol. *12*, 643–662.

Talairach, J., and Tournoux, P. (1988). A Co-Planar Stereotaxic Atlas of the Human Brain (New York: Thieme).

Thomson, J.J. (1986). Rights, Restitution, and Risk: Essays, in Moral Theory (Cambridge, MA: Harvard University Press).

Wager, T.D., and Smith, E.E. (2003). Neuroimaging studies of working memory: A meta-analysis. Cogn. Affect. Behav. Neurosci. *3*, 255–274.

Wager, T.D., Rilling, J.K., Smith, E.E., Sokolik, A., Casey, K.L., Davidson, R.J., Kosslyn, S.M., Rose, R.M., and Cohen, J.D. (2004). Placebo-induced changes in fmri in the anticipation and experience of pain. Science *303*, 1162–1167.

Woods, R.P., Cherry, S.R., and Mazziotta, J.C. (1992). Rapid automated algorithm for aligning and reslicing pet images. J. Comput. Assist. Tomogr. *16*, 620-633.