Judgment before principle: engagement of the frontoparietal control network in condemning harms of omission

Fiery Cushman, Dylan Murray, Shauna Gordon-McKeon, Sophie Wharton, and Joshua D. Greene

Department of Cognitive, Linguistic and Psychological Sciences, Box 1821, Brown University, Providence, RI 02912, USA

Ordinary people make moral judgments that are consistent with philosophical and legal principles. Do those judgments derive from the controlled application of principles, or do the principles derive from automatic judgments? As a case study, we explore the tendency to judge harmful actions morally worse than harmful omissions (the 'omission effect') using fMRI. Because ordinary people readily and spontaneously articulate this moral distinction it has been suggested that principled reasoning may drive subsequent judgments. If so, people who exhibit the largest omission effect should exhibit the greatest activation in regions associated with controlled cognition. Yet, we observed the opposite relationship: activation in the frontoparietal control network was associated with condemning harmful omissions—that is, with *overriding* the omission effect. These data suggest that the omission effect arises automatically, without the application of controlled cognition. However, controlled cognition is apparently used to overcome automatic judgment processes in order to condemn harmful omissions.

Human morality is characterized by certain widely shared patterns of judgment. Among the most studied is the distinction between active and passive harm, which we refer to as the 'omission effect'. Most people judge it morally worse to actively harm a person than to passively allow a preventable harm to a person (Spranca *et al.*, 1991; Ritov and Baron, 1999; Baron and Ritov, 2004; Cushman *et al.*, 2006). For instance, it is judged morally worse for a physician to actively administer a lethal injection to a terminally ill patient than to passively withhold available lifesaving treatments, omitting to save the patient. This distinction has been endorsed by the US Supreme Court (Vacco v. Quill, 1997), the American Medical Association (Rachels, 1975), and by many moral philosophers (Fischer and Ravizza, 1992). But why is killing judged to be worse than letting die?

We consider two explanations. According to the 'rule hypothesis', people have an explicit rule discriminating active and passive harm that they apply to particular cases, much as a judge applies explicit legal rules. If the rule hypothesis is correct, then people who exhibit an omission effect are doing so by engaging in controlled, effortful cognition—retrieving a rule and applying it.

According to the 'automaticity hypothesis', people have automatic mechanisms of moral judgment that are more sensitive to active harm than the failure to prevent harm, much as a jury might be more swayed by lively direct testimony than by a dry written deposition. If the automaticity hypothesis is correct, then exhibiting an omission effect does not require any controlled, effortful cognition. To the contrary, *overriding* the omission effect would require controlled, effortful cognition.

These hypotheses about the omission effect draw on a basic distinction between automatic and controlled psychological processes that lie at the heart of current research in moral psychology (Greene *et al.*, 2001; Haidt, 2001; Pizarro and Bloom, 2003; Cushman, *et al.*, 2006; Bartels, 2008; Lombrozo, 2009; Cushman *et al.*, in press). They target a fundamental chicken-and-egg problem: Which comes first, judgments of particular cases, or the principles governing them? That is, do people's judgments conform to moral principles because they have been exposed to such principles, directly or indirectly? Or, do the principles merely summarize a preexisting pattern in intuitive moral judgment (Haidt, 2001)? We use the omission effect as a case study of this general question. Because the rule hypothesis predicts greater controlled cognitive processing in those who exhibit an omission effect, while the automaticity hypothesis predicts greater controlled cognitive processing in those who override the omission effect, functional neuroimaging provides a useful method to distinguish between these rival hypotheses.

The rule hypothesis

Past studies of the omission effect are consistent with the rule hypothesis. In one study, participants judged contrasting cases of active *vs* passive harm and were asked to justify any discrepancies in their responses to the two types of cases (Cushman *et al.*, 2006). A large majority of participants explained their discrepant judgments by appealing to a moral distinction between active and passive harm. Although participants' explicit justifications may have been *post hoc* rationalizations of prior intuitive moral judgments (Haidt, 2001), their ability to reference an explicit principle is at least consistent with the rule hypothesis.

Evidence from functional neuroimaging has also been interpreted to support the rule hypothesis (Borg *et al.*, 2006). Cases contrasting active and passive harms evoked greater activation in a network of brain regions associated with explicit reasoning compared with cases contrasting two active harms. This effect was particularly strong in a rostral region of dorsolateral prefrontal cortex (DLPFC). The authors concluded that the omission effect is partially mediated by explicit reasoning. However, their study did not correlate levels of controlled cognition with the size of the omission effect, the key analysis employed in this study. The DLPFC activity observed might reflect participants' efforts to judge in a manner consistent with an explicit rule condemning harmful actions more than omissions. Alternatively, however, it might reflect participants' efforts to *override* an automatic tendency to judge harmful actions more harshly than omissions. Our study directly assesses these rival explanations.

The automaticity hypothesis

The automaticity hypothesis proposes that automatic mechanisms used to comprehend and judge others' behaviors are more responsive to actions than to omissions. If so, then the condemnation of a harmful omission would require extra controlled cognition, above-and-beyond what is required for the condemnation of harmful actions. At a very general level, this is consistent with evidence that processing negations (e.g., 'he did *not* administer the treatment') and counterfactual dependencies (e.g., 'if he had administered the treatment, then the patient would be alive') engages controlled, effortful cognition (Wason, 1968; Carpenter and Just, 1975; Smith and DeCoster, 2000; Deutsch *et al.*, 2006; Reverberi *et al*, 2009).

Possibly, the direct and active causation of harm triggers an automatic process of moral judgment, while reasoning about omissions requires cognitive control to process counterfactual responsibility for harm. Indeed, past research suggests that humans evaluate the moral status of harmful actions by at least two distinct pathways: an automatic process that responds to a select sub-class of especially 'direct' or 'personal' harms, and a controlled process that evaluates the permissibility of actions given their foreseen consequences (reviewed in Cushman *et al.*, 2010; Greene, in press).

Experimental hypotheses and design

This study tests divergent predictions of the rule and automaticity hypotheses using functional neuroimaging. It depends on an analysis of individual differences. The rule hypothesis predicts that people who consistently judge actions to be morally worse than omissions will show extra processing in regions associated with controlled cognition compared to people who judge actions and omissions identically. This processing would reflect the application of an explicit rule distinguishing active from passive harm. The automaticity hypothesis predicts exactly the opposite effect: people who judge actions and omissions identically should show extra processing in regions associated with controlled cognition compared to people who judge actions to be morally worse than omissions. This processing would reflect the use of controlled processes to condemn harmful omissions.

The frontoparietal control network, and particularly the dorsolateral prefrontal cortex, are key regions of interest that indicate engagement of controlled cognition. The frontoparietal control network is a network of brain regions preferentially engaged during explicit reasoning tasks and tasks that require overriding automatic responses (Miller and Cohen, 2001; Vincent *et al.*, 2008). It includes regions of dorsal medial prefrontal cortex (DMPFC), anterior cingulate cortex (ACC), dorsolateral prefrontal cortex (DLPFC), and anterior inferior parietal lobule (aIPL). Among these regions, a wealth of evidence indicates that the DLPFC (Brodmann areas (BA) 9 and 10) plays a critical role in controlled reasoning processes, and especially in the selection of appropriate rules to apply to a problem (Goel *et al.*, 1997; Wagner *et al.*, 2001; Bunge, 2004; Goel and Dolan, 2004; Boettiger and D'Esposito, 2005; Canessa *et al.*, 2005). Activity in the DLPFC is also associated with the controlled application of utilitarian principles to the judgment of moral dilemmas (Greene *et al.*, 2004). And, as noted above, this region showed increased activity for trials contrasting active *vs* passive harms compared to trials contrasting active harms in a previous study of the omission effect (Borg *et al.*, 2006).

Patterns of intrinsic functional connectivity during a resting state can be used to identify components of the frontoparietal control network (Dosenbach *et al.*, 2007; Seeley *et al.*, 2007; Vincent *et al.*, 2008). This approach has been used as a basis for comparison with task-related activations (Wig *et al.*, 2009; Spreng *et al.*, 2010). We adopt the same approach here. Our scanning protocol included a resting state functional scan, and we used patterns of functional connectivity derived from that scan to identify the frontoparietal control network and compare it with patterns of activation obtained during moral judgment.

Participants were presented with a series of 24 moral scenarios, comprising 12 cases of actions leading to harm and 12 cases of omissions leading to harm. In all cases, the harmdoer was fully aware of the harmful consequences of his or her choice. Our full stimulus set comprised 48 scenarios

arranged into 24 matched action/omission pairs, but each participant viewed only one item from each matched pair. Our goal was to assess the cognitive processes engaged in evaluating a single moral event (action or omission) as often happens in a natural context, rather than to assess the cognitive processes engaged in an explicit comparison between matched events, which is unusual outside of specialized settings such as the laboratory. Our design allowed us to do so by precluding participants from making an explicit comparison between matched versions of the same scenario, which would inevitably have highlighted the action/omission distinction.

Each of the moral scenarios used in this study contains some source of conflict between competing interests, but the scenarios differ in the nature of the interests involved and in whose interests they are. Most of the scenarios describe an agent who must choose between harming one person and the welfare of several others, but some involve a different competing interest, such as the agent's own welfare. Some of the scenarios presented tradeoffs in lives saved *vs* lost, but others involved interests of a different nature: physical harms short of death, emotional harms such as the loss of valuable objects, or personal ambitions such as the desire to win a competition. In this sense the stimulus set was fairly heterogeneous. A disadvantage of this heterogeneity is that it limits our interpretation of trends that apply across both action and omission cases. The focus of our study, however, is a comparison between action and omission cases, and heterogeneity was strictly controlled across each action/omission pair: any unique factor present in an action case was precisely matched in the corresponding omission case (e.g., if an action case involved a choice between another's life and one's own personal ambitions, then the corresponding omission case also involved a choice between the very same life and personal ambitions). Moreover, the heterogeneity of the action/omission pairs (compared to each other) affords an advantage: any differences we find between judgments concerning action cases *vs* omission cases apparently apply across a diverse range of situations and tradeoffs.

Scenarios were presented on a series of three screens (Figure 1): a *background* screen establishing general context, an *options* screen indicating the agent's choices and their consequences, and a *behavior* screen indicating the agent's chosen behavior. Participants were then asked to judge the moral wrongness of the agent's behavior on a five-point Likert scale (a *response* screen). Past research shows that participants spontaneously process information relevant to a moral judgment as soon as that information is presented (Young and Saxe, 2008). In our study, the *background* screen is identical across action and omission variants and is therefore not of interest. The *options* screen differs between conditions in that a focal act of harm can optionally be caused either by an action or by omission of action. Thus, activation of regions associated with deliberation and conflict between competing options is expected during this period. Of

greatest interest, though, is the *behavior* period. Because the agent is first clearly identified as either performing an action or an omission on the behavior screen, we expect neural correlates of moral judgment to be most strongly activated, and also for any differences between conditions to be greatest, in this period. These activations may persist through the beginning of the response period, but as we report below subjects typically indicated a response fewer than three seconds from the onset of this screen.

METHODS

Thirty five right-handed participants participated in the study for payment (age 18–28, 20 female). All participants were native English speakers, had normal or corrected-to-normal vision, and gave written informed consent in accordance with the requirements of the Internal Review Board at Harvard University. Participants were scanned at 3T (at the Harvard Brain Sciences Center in Cambridge, MA). An initial battery of scans included a 1-mm^3 resolution anatomical scan and a 6.25 min functional resting state scan. During the subsequent moral judgment task standard echoplanar imaging procedures were used (TR = 2.5 s, TE = 30 ms, flip angle = 85°, 39 3-mm-thick slices parallel to the AC-PC line).

Stimuli consisted of 24 moral scenarios, each written in a separate 'action' and 'omission' version (for a sample scenario, see Figure 1; the full text of all scenarios is available in Supplementary Materials). Participants saw either the action or omission version of each scenario, for a total of 12 action trials and 12 omission trials, counterbalanced between participants and separated into three runs of eight scenarios each. Scenarios were presented on a series of three screens that advanced automatically (Figure 1): a *background* screen establishing general context (12 s), an *options* screen indicating the agent's choices and their consequences (14 s), and a *behavior* screen indicating the agent's chosen behavior (7 s). Participants were then asked to judge the moral wrongness
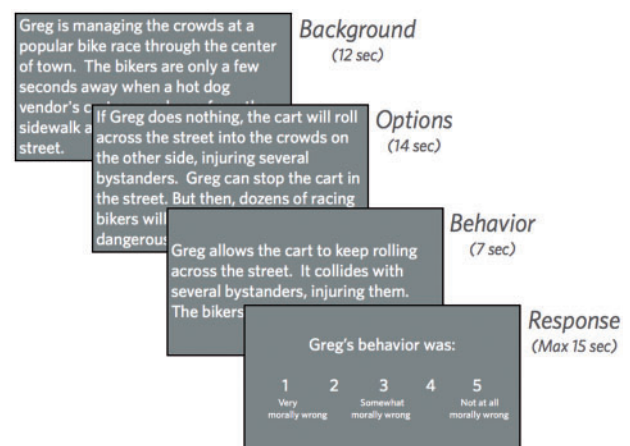
Fig. 1 Schematic representation of stimulus presentation and response for a sample omission trial.
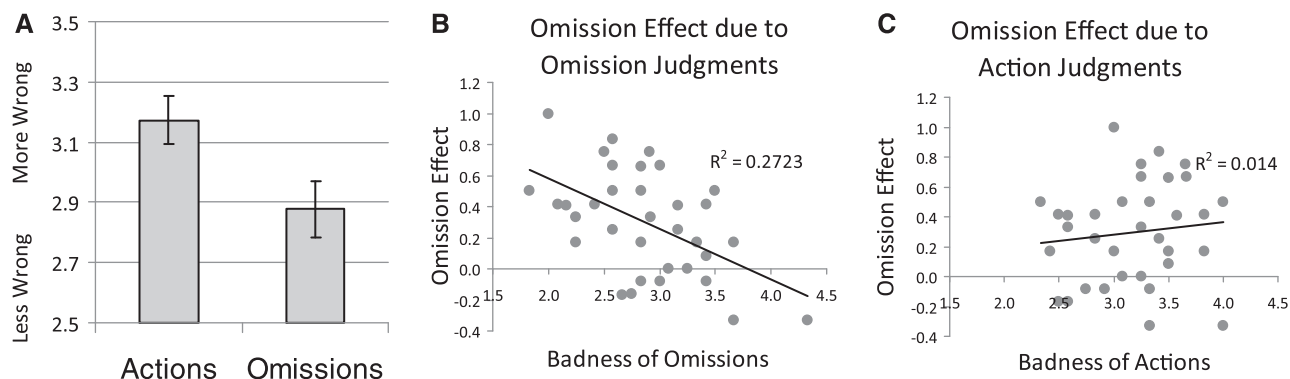
Fig. 2 (**A**) Mean judgment by trial type (action *vs* omission), showing a reliable omission effect. The size of each participant's omission effect was predicted by (**B**) the mean judgment of omission trials, but not by (**C**) the mean judgment of action trials.

of the agent's behavior on a five-point Likert scale (a *response* screen) within 15 seconds, using a five-button box held in the right hand. Fixation blocks (10 seconds) were interleaved between each story, plus the balance of remaining response time (given typical response times reported below, the total interstimulus interval averaged about 17 seconds). The text of the scenarios was presented in white on a black background. Scenarios were projected onto a screen via Matlab 2009a running on a MacBook Pro laptop.

fMRI data were analyzed using SPM8 (http://www.fil.ion. ucl.ac.uk/spm). Each participant's data were motion corrected and normalized onto a common brain space (Montreal Neurological Institute template) using default procedures. Data were smoothed using a Gaussian filter (full width half maximum = 6 mm). The Artifact Detection Tool (gablab.mit.edu/software) was used to regress out individual timepoints with contrast values greater than three standard deviations from the mean or with motion parameters exceeding 0.6 mm in any direction. Participants with more than 50/558 timepoints above threshold were excluded from group analysis ($N = 6$). The experiment used a long event related design and was modeled using SPM's canonical hemodynamic response. Each screen (background, options, behavior, and response) was modeled as a separate event, and action and omission trials were modeled separately.

The correlations reported in the results section and Figure 4 are derived from a regression model predicting the behavioral omission effect (average action–average omission judgment made at the *response* screen) for each participant from the average β associated with a given contrast across all voxels in a region of interest (obtained using the MARSBAR toolbox, marsbar.sourceforge.net). The omission effects used for brain–behavior correlations were calculated using only trials entered in the fMRI analysis (for three participants, some trials were dropped due to computer malfunction).

Functional connectivity analyses were performed with CONN (web.mit.edu/swg/software). Resting state scans

were preprocessed as described above. Further preprocessing conformed to CONN defaults and removed variance predicted by BOLD signal detected in white matter and cerebrospinal fluid, realignment parameters and their first-order derivatives, and global signal. A high band pass filter of 0.08 Hz was also applied.

RESULTS AND DISCUSSION
Analysis of behavioral results
Participants judged active harms to be significantly morally worse than passive harms $t(33) = 5.1$, $P < .0001$ (Figure 2A). The size of the omission effect for each individual was calculated by subtracting the mean judgment of omission cases from the mean judgment of action cases on the 1–5 scale of moral wrongness ($M = 0.30$, $\sigma = 0.34$). Individual differences in the size of the omission effect were strongly predicted by participants' judgments of harmful omissions $r = -0.52$, $N = 34$, $P < 0.005$ (Figure 2B), but not by participants' judgments of harmful actions $r = 0.12$, $N = 34$, $P = 0.50$ (Figure 2C). Specifically, the more harshly participants judged omissions, the less they exhibited a difference in the judgment of actions and omissions.

Thus, behavioral results suggest that meaningful differences in the omission effect were driven by the way that people responded to harmful omissions, rather than the way that people responded to harmful actions. From the perspective of the rule hypothesis, this suggests individual differences in the application of an explicit rule exonerating harmful omissions. From the perspective of the automaticity hypothesis, this suggests individual differences in the controlled processing of harmful omissions leading to their condemnation.

Average response times to action trials ($M = 2.6$ s) were slightly shorter than to omission trials ($M = 2.9$ s), and this difference was marginally significant $t(31) = 1.9$, $P < 0.07$. There was no significant relationship between response time and moral judgment.

fMRI analysis of action vs omission trials

A whole-brain random-effects analysis compared the neural activity for action *vs* omission trials. Separate analyses were performed for the second block of text (*options*), which presented the behavioral options available to the agent and the third block of the text (*behavior*), which described the ultimate behavioral choice of the agent. We applied a voxel-level statistical threshold of $P < 0.001$ and a cluster threshold of $k = 20$, ensuring a minimum cluster-level significance of $P \leq 0.051$ uncorrected for multiple comparisons. Below we note which clusters are not significant when corrected for multiple comparisons to achieve a family-wise error rate < 0.05.

Options period

For the second block of text (*options*) one cluster of 33 voxels in the rostral anterior cingulated cortex (rACC) survived this statistical threshold for the actions > omission contrast (Figure 3; MNI peak 3, 32, 7, cluster level $P = 0.014$ uncorrected); however, this cluster was not statistically significant
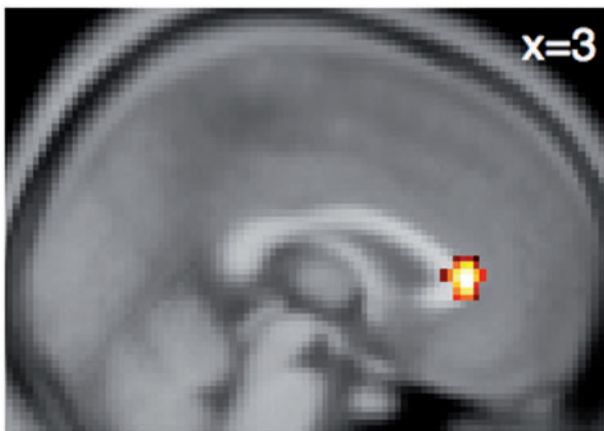
Fig. 3 Region identified by the contrast actions > omissions for the *options* block of text. Voxel-wise $P < 0.001$, $k \geq 20$.

after correction for multiple comparisons and so must be interpreted with caution. There were no significant activations for the reverse contrast.

Behavior period

For the third block of text (*behavior*), no clusters survived the statistical threshold for the actions > omissions contrast. However, the omissions > actions contrast identified nine distinct clusters comprising 1549 voxels (Table 1 and Figure 3). These included bilateral DLPFC activations spanning BA 6, 8, 9, and 10, consistent with the increased engagement of controlled cognitive processes for judgments of harms of omission. It also included activations in medial prefrontal cortex (MPFC), posterior cingulate (PC), bilateral inferior parietal lobule (IPL), and bilateral middle temporal gyrus (MTG). The cluster-level significance values for right and left MTG do not individually survive correction for multiple comparisons, however, and so must be interpreted with caution.

There is apparent overlap between the neural regions identified above and the frontoparietal control network. In order to visualize this overlap, we localized the control network in each participant using an analysis of functional connectivity during a 6.25 min resting-state scan administered prior to the moral judgment task. Figure 4 compares group-level maps of the control networks with the regions identified in the group-level omission > action contrast for the *behavior* block of the text. Specifically, Figure 4 indicates voxels exhibiting functional correlations of $P > 0.001$ with bilateral seed regions of anterior IPL (in blue). Our seed regions were modeled after a recent study by Vincent *et al.* (2008), and comprised spheres of approximately $2.2\,cm^2$ volume centered on -52, -49, 47 and 52, -46, 46. We also used this statistical map as the basis of a small volume analysis of the omission > action contrast for the behavior period, and the results of this supplementary analysis are presented in Table 1.

Table 1. Regions identified by the omission > action contrast for the *behavior* block of text applying a voxel-wise significance threshold of $P < 0.001$ uncorrected and a minimum cluster size of 20

Region	Brodmann Areas	MNI peak	Whole brain analysis		Peak z	Small Volume Analysis	
			Cluster size	Cluster significance, *P*-value		Cluster size	Cluster significance
A. rDLPFC (BA 10)	10	(24, 56, 16)	161	<0.001	4.26	83	<0.001
B. MPFC	6, 8, 9, 32	(6, 41, 40)	501	<0.001	4.19	341	<0.001
C. rDLPFC (BA 9)	9	(39, 23, 37)	102	<0.001	4.2	100	<0.001
D. lDLPFC (BA 8)	8	(−45, 8, 43)	135	<0.001	4.17	103	<0.001
E. lMTG*	21	(−60, −31, −8)	23	0.039*	3.97	23	0.039*
F. rMTG*	21	(57, −34, −8)	39	0.01*	4.62	23	0.039*
G. PC	7, 31	(−6, −43, 40)	327	<0.001	3.93	203	<0.001
H. lIPS	40	(−51, −52, 31)	114	<0.001	4.48	114	<0.001
I. rIPS	40	(51, −61, 37)	147	<0.001	4.32	146	<0.001

*Indicates cluster-level significance values of $P > 0.05$ when corrected for multiple comparisons.
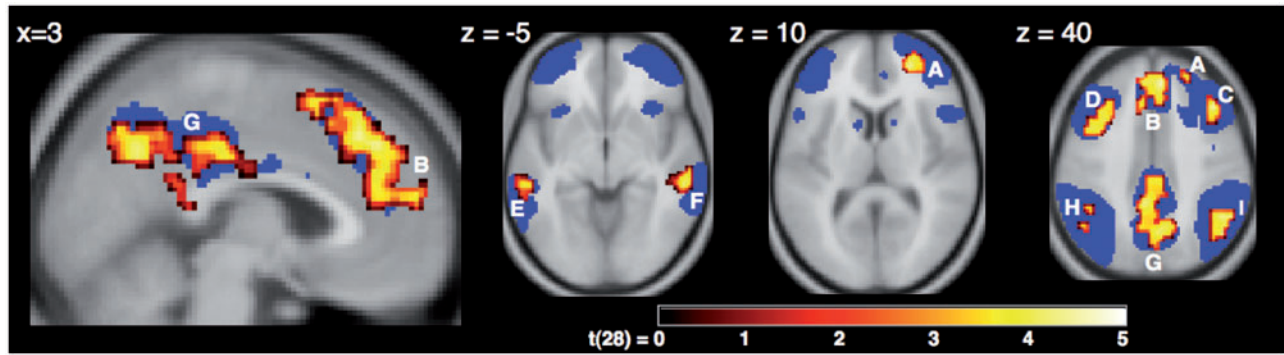
Fig. 4 Heat map indicates regions identified by the omission > action contrast for the *behavior* block of text. Voxel-wise $P < 0.001$, $k > 20$. Area labels correspond to Table 1. Results are overlaid on a representation of the frontoparietal control network derived via functional connectivity during resting state, with blue indicating voxels correlating at $P < 0.001$ with bilateral aIPL.

Among the regions preferentially engaged during omission trials was the DLPFC, the region most closely associated with controlled cognition (Miller, 2000). From the perspective of the rule hypothesis, this is consistent with the application of an explicit rule to cases of harmful omissions. From the perspective of the automaticity hypothesis, this is consistent with the engagement of controlled processes to understand and interpret the omission of an action. An analysis of individual differences, however, allows us to select between these rival hypotheses.

Response period

During the *response* period there were no activations above threshold for either the actions > omissions contrast or the omissions > actions contrast.

Analysis of individual differences

Our analysis of individual differences begins with ROIs identified by the omissions > actions contrast for the *behavior* block of the text. Of particular interest are individual differences in activation in regions of DLPFC: Is increased BOLD response during omission trials associated with an increased omission effect (as predicted by the rule hypothesis), or with a decreased omission effect (as predicted by the automaticity hypothesis)?

For each of the nine ROIs we separately regressed the size of the neural omission > action contrast during the *behavior* block of text on the size of each participant's behavioral omission effect. All nine ROIs exhibited the same directional effect: the *larger* the BOLD activation for omission trials relative to action trials, the *smaller* the behavioral omission effect. The average regression coefficient across regions differed significantly from zero as assessed by a one-sample *t*-test $M = -0.22$, $t(8) = -5.0$, $P = 0.001$, indicating that the uniform directional effect observed across regions is unlikely to be due to chance. This finding supports the automaticity hypothesis: the more a person judges harmful omissions on parity with harmful actions, the more they engage cognitive control during the judgment of omissions. These
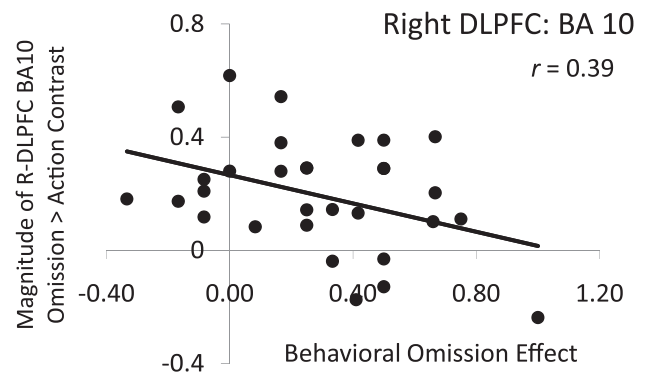
Fig. 5 Correlation between the behavioral omission effect (mean action judgment—mean omission judgment) and differential BOLD response to action vs omission trials within rDLPFC. The greater the BOLD response to omission trials relative to action trials, the lesser the magnitude of the omission effect.

effects were statistically significant (two-tailed) within two ROIs, the right DLPFC, BA 10 $r = -0.39$, $t = -2.19$, $P = 0.04$ (Figure 5) and the right MFG $r = -0.37$, $t = -2.06$, $P < 0.05$. The relationship between BOLD response and behavior that we observed is noteworthy. It is precisely when no behavioral differences are observed that the associated patterns of neural response differ most greatly.

GENERAL DISCUSSION

Consideration of harmful omissions, compared to harmful actions, preferentially engaged multiple neural regions within the frontoparietal control network, as defined by analyses of functional connectivity during rest. Across these regions, participants who exhibited the *largest* increases in fMRI BOLD signal during omission trials tended to exhibit the *smallest* differences in moral judgments of harmful actions *vs* omissions (i.e., exhibited the smallest omission effects). This relationship was statistically significant in right DLPFC, suggesting a role for controlled cognition in the elimination of the omission effect.

Despite its *prima facie* appeal, and contrary to previous suggestions (Borg *et al.*, 2006; Cushman *et al.*, 2006), the rule hypothesis is not well supported by the present data. Our behavioral results show that individual differences in the size of the omission effect are driven by the judgment of omissions in particular. Thus, according to the rule hypothesis, participants who condemn omissions harshly fail to consciously apply the relevant rule, and should therefore exhibit reduced control network activity. Yet, our data reveal precisely the opposite pattern: participants who most strongly condemned omissions exhibited increased engagement of DLPFC, and the control network more broadly, during omission trials. In other words, controlled cognition is associated not with conforming to the omission effect but with overriding it.

Our data thus support the automaticity hypothesis instead. When participants engaged similar cognitive processes for action and omission trials, they tended to judge actions more harshly. This accords with a growing body of evidence that a select class of canonically harmful behaviors—ones performed actively, intentionally and in a physically direct manner—automatically elicit moral condemnation (Greene *et al.*, 2001, 2004; Cushman *et al.*, 2006; Greene *et al.*, 2008; Greene *et al.*, 2009).

Our analysis identified one region, rACC, that responded more strongly to action trials than to omission trials, but the extent of this activation (33 voxels) was not sufficient to survive a multiple comparisons correction. While this activation must be interpreted with caution, then, it may be worth speculating why we find it. Research indicates that activity in rACC is associated with emotional conflict (Bush *et al.*, 2000; Mohanty *et al.*, 2007). For instance, it shows increased activation for incongruent trials on an 'emotional' stroop task that pairs fearful faces with happy words (or *vice versa*) compared with congruent trials where the faces and the words have the same emotional valence (Etkin *et al.*, 2006). As noted above, the moral scenarios used in this study each contain some source of emotional conflict: most present a choice between harm to one person and the welfare of several others, while some others involve another competing interest such as one's own welfare. Thus, activation in rACC may have been stronger for action trials than for omission trials because the automatic condemnation elicited by harmful actions evokes stronger emotional conflict. However, because of the heterogeneity of the competing interests in our scenarios (e.g., self-interest *vs* other-interest), the present experiment is not ideally suited to investigating the neural correlates of emotional conflict and our interpretation of rACC activation remains tentative.

Beyond their particular application to the omission effect, our results speak more generally to the role of automatic *vs* controlled processes in moral judgment. As noted above, ordinary people make moral judgments that are consistent with legal and philosophical principles, but it is not clear whether the principles cause those judgments or the judgments cause the principles. Because past research shows that most people are able to articulate a principled basis for the omission effect (Spranca *et al.*, 1991; Cushman *et al.*, 2006) it is a likely candidate for the application of general rules to specific cases. Our findings suggest the opposite conclusion; however, and thus contribute to a growing body of evidence that legal and philosophical principles often formalize basic properties of the automatic, intuitive moral judgments of ordinary people (Robinson and Darley, 1995; Mikhail, 2002, 2009; Cushman, 2008; Greene, 2008; Cushman and Greene, in press).

Correspondence between explicit moral principles and automatic moral judgments raises the important question of how these distinct processes are integrated. For the purposes of our experiment it is convenient to contrast the roles of automatic and controlled processing, but in many contexts of moral judgment-making these processes will exert compatible influences and may interface in mutually supportive roles. In addition, the generalization of moral principles from patterns of intuitive response demands some interface between automatic and controlled processes. Thus, the integration of automatic and controlled processing—in moral judgment and other judgments—is an important direction for further research (for additional discussion, see Moll *et al.*, 2008; Frank *et al.*, 2009; Cushman *et al.*, 2010).

While our findings support the automaticity hypothesis generally, they contradict the strongest version of that hypothesis, according to which controlled cognitive processes rarely contribute to ordinary moral judgment, especially in the absence of conflicting intuitions (Haidt, 2001). To the contrary, we find that controlled processing plays a key role in overriding the omission effect. As such, our data support the view that both automatic and controlled processes make meaningful, dissociable contributions to moral judgment (Pizarro and Bloom, 2003; Cushman *et al.*, 2006; Greene, 2008, in press).

SUPPLEMENTARY DATA

Supplementary data are available at *SCAN* online.

Conflict of Interest

None declared.

REFERENCES

Baron, J., Ritov, I. (2004). Omission Bias, individual differences, and normality. *Organizational Behavior and Human Decision Processes*, 94, 74–85.

Bartels, D. (2008). Principled moral sentiment and the flexibility of moral judgment and decision making. *Cognition*, 108(2), 381–417.

Boettiger, C., D'Esposito, M. (2005). Frontal networks for learning and executing arbitrary stimulus-response associations. *Journal of Neuroscience*, 25(10), 2723.

Borg, J.S., Hynes, C., Van Horn, J., Grafton, S.T., Sinnott-Armstrong, W. (2006). Consequences, action, and intention as factors in moral judgments: an fMRI investigation. *Journal of Cognitive Neuroscience*, 18(5), 803–837.

Bunge, S. (2004). How we use rules to select actions: a review of evidence from cognitive neuroscience. *Cognitive, Affective, and Behavioral Neuroscience*, 4(4), 564.

Bush, G., Luu, P., Posner, M.I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Sciences*, 4(6), 215–222.

Canessa, N., Gorini, A., Cappa, S., et al. (2005). The effect of social content on deductive reasoning: an fMRI study. *Human brain mapping*, 26(1), 30–43.

Carpenter, P., Just, M. (1975). Sentence comprehension: a psycholinguistic processing model of verification. *Psychological Review*, 82(1), 45–73.

Cushman, F.A. (2008). *The Origins of Moral Principles*. Unpublished Doctoral Dissertation. Cambridge: Harvard University.

Cushman, F.A., Greene, J.D. (2011). The philosopher in the theater. In: Mikulincer, M., Shaver, P.R., editors. *Social Psychology of Morality: The Origins of Good and Evil*. New York: APA Press.

Cushman, F.A., Young, L., Greene, J.D. (2010). Multi-system moral psychologyDoris J. M., T. M. P. R. Group editors. *The Oxford Handbook of Moral Psychology*. New York: Oxford University Press.

Cushman, F.A., Young, L., Greene, J.D. (2010). Multi-system moral psychology. *The Oxford Handbook of Moral Psychology*. Oxford: Oxford University Press.

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

Deutsch, R., Gawronski, B., Strack, F. (2006). At the boundaries of automaticity: negation as reflective operation. *Journal of Personality and Social Psychology*, 91(3), 385.

Dosenbach, N.U.F., Fair, D.A., Miezin, F.M., et al. (2007). Distinct brain networks for adaptive and stable task control in humans. *Proceedings of the National Academy of Sciences*, 104(26), 11073.

Etkin, A., Egner, T., Peraza, D.M., Kandel, E.R., Hirsch, J. (2006). Resolving emotional conflict: a role for the rostral anterior cingulate cortex in modulating activity in the amygdala. *Neuron*, 51(6), 871–882.

Fischer, J.M., Ravizza, M. (1992). *Ethics: Problems and Principles*. New York: Holt, Rinehart and Winston.

Frank, M.J., Cohen, M.X., Sanfey, A.G. (2009). Multiple systems in decision making. *Current Directions in Psychological Science*, 18(2), 73.

Goel, V., Dolan, R. (2004). Differential involvement of left prefrontal cortexin inductive and deductive reasoning. *Cognition*, 93(3), B109–21.

Goel, V., Gold, B., Kapur, S., Houle, S. (1997). The seats of reason? An imaging study of deductive and inductive reasoning. *Neuroreport*, 8(5), 1305.

Greene, J.D. (2008). The secret joke of Kant's soul. In: Sinnott-Armstrong, W., editor. *Moral Psychology*, Vol. 3, Cambridge, MA: MIT Press.

Greene, J.D. (in press). The moral brain and how to use it. Penguin Group, New York.

Greene, J.D., Cushman, F.A., Stewart, L.E., Lowenberg, K., Nystrom, L.E., Cohen, J.D. (2009). Pushing moral buttons: the interaction between personal force and intention in moral judgment. *Cognition*, 111(3), 364–371.

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

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

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

Lombrozo, T. (2009). The role of moral commitments in moral judgment. *Cognitive Science*, 33, 273–286.

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

Mikhail, J. (2009). Moral grammar and intuitive jurisprudence: a formal model of unconscious moral and legal knowledge. *Psychology of learning and motivation*, 50, 27–100.

Miller, E.K. (2000). The prefrontal cortex and cognitive control. *Nature Reviews*, 1, 59–65.

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

Mohanty, A., Engels, A.S., Herrington, J.D., et al. (2007). Differential engagement of anterior cingulate cortex subdivisions for cognitive and emotional function. *Psychophysiology*, 44(3), 343–51.

Moll, J., De Oliveira Souza, R., Zahn, R. (2008). The neural basis of moral cognition. *Annals of the New York Academy of Sciences*, 1124(1), 161–80.

Pizarro, D.A., Bloom, P. (2003). The intelligence of the moral intuitions: comment on Haidt (2001). *Psychological Review*, 110(1), 193–6; discussion 197–8.

Rachels, J. (1975). Active and passive euthanasia. *New England Journal of Medicine*, 292, 78–80.

Reverberi, C., Burigo, M., Cherubini, P. (2009). A subliminal premise can automatically trigger an elementary deductive inference. *CogSci 2009 Proceedings*.

Ritov, I.I., Baron, J. (1999). Protected Values and Omission Bias. *Organizational Behavior and Human Decision Processes*, 79(2), 79–94.

Robinson, P.H., Darley, J.M. (1995). *Justice, Liability and Blame*. Boulder: Westview Press.

Seeley, W.W., Menon, V., Schatzberg, A.F., et al. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *The Journal of Neuroscience*, 27(9), 2349.

Smith, E., DeCoster, J. (2000). Dual-process models in social and cognitive psychology: conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, 4(2), 108.

Spranca, M., Minsk, E., Baron, J. (1991). Omission and commission in judgment and choice. *Journal of Experimental Social Psychology*, 27(1), 76–105.

Spreng, R.N., Stevens, W.D., Chamberlain, J.P., Gilmore, A.W., Schacter, D.L. (2010). Default network activity, coupled with the frontoparietal control network, supports goal-directed cognition. *Neuroimage*, 53(1), 303–317.

Vacco v. Quill, 521 U.S. 793 (U.S. Supreme Court 1997).

Vincent, J., Kahn, I., Snyder, A., Raichle, M., Buckner, R. (2008). Evidence for a frontoparietal control system revealed by intrinsic functional connectivity. *Journal of Neurophysiology*, 100(6), 3328.

Wagner, A., Maril, A., Bjork, R., Schacter, D. (2001). Prefrontal contributions to executive control: fMRI evidence for functional distinctions within lateral prefrontal cortex. *Neuroimage*, 14(6), 1337–1347.

Wason, P. (1968). Reasoning about a rule. *The Quarterly Journal of Experimental Psychology*, 20(3), 273–281.

Wig, G.S., Buckner, R.L., Schacter, D.L. (2009). Repetition priming influences distinct brain systems: evidence from task-evoked data and resting-state correlations. *Journal of Neurophysiology*, 101(5), 2632.

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