nature neuroscience

Corticolimbic gating of emotion-driven punishment

Michael T Treadway[1,2,10], Joshua W Buckholtz[3,10], Justin W Martin[3], Katharine Jan[4], Christopher L Asplund[5], Matthew R Ginther[6], Owen D Jones[6–8] & René Marois[9]

Determining the appropriate punishment for a norm violation requires consideration of both the perpetrator's state of mind (for example, purposeful or blameless) and the strong emotions elicited by the harm caused by their actions. It has been hypothesized that such affective responses serve as a heuristic that determines appropriate punishment. However, an actor's mental state often trumps the effect of emotions, as unintended harms may go unpunished, regardless of their magnitude. Using fMRI, we found that emotionally graphic descriptions of harmful acts amplify punishment severity, boost amygdala activity and strengthen amygdala connectivity with lateral prefrontal regions involved in punishment decision-making. However, this was only observed when the actor's harm was intentional; when harm was unintended, a temporoparietal-medial-prefrontal circuit suppressed amygdala activity and the effect of graphic descriptions on punishment was abolished. These results reveal the brain mechanisms by which evaluation of a transgressor's mental state gates our emotional urges to punish.

When determining how best to punish the harmful actions of another, individuals frequently rely on emotional heuristics to guide their decisions[1,2]. Intuitive, gut responses to the damage done have been hypothesized to serve as a form of internal emotional 'evidence' that is used to select the suitable just desserts[3–5]. Justice, however, requires that punishment take into account not only the negative emotions elicited by harm, but also an evaluation of the transgressor's intent.

For example, even the most severe harms do not warrant punishment when they occur accidentally and without negligence, such as when a pedestrian fatality results from the unforeseeable mechanical failure of a new car's brakes[4]. An actor's mental state—whether it is purposeful, knowing, reckless, negligent or blameless[6]—can markedly affect how severely he or she is punished for the harm committed[7–9]. Thus, despite the powerful role that emotional responses to harm have in punishment decision-making, an actor's mental state can outweigh harm-based affective signals in determining an appropriate sanction. Indeed, the ability of mental state information to overrule harm is a foundational principle of modern criminal justice systems[1,10–12]. However, although research has begun to reveal the brain mechanisms by which we can decode the mental state of others[13–15], affectively react to harm[16] and make punishment decisions[15,17–20], the neural dynamics through which representations of an actor's mental state can gate harm-dependent emotional heuristics are currently unknown.

We sought to elucidate the neural circuitry through which a transgressor's mental state can modulate harm-based affective responses during punishment decision-making. To that end, we independently manipulated actor blameworthiness and affective content across a series of text-based scenarios that described various levels of harms committed by the actor. To manipulate blameworthiness, half of the scenarios, those in the 'intentional' condition, depicted a protagonist who explicitly desired to cause the harm that actually occurred (purposeful in legal terms). In the other half of the scenarios, those in the 'unintentional' condition, the protagonist caused identical harms, but without any blameworthy intent to have done so. Scenarios varied in terms of severity of harm from property damage to death, with all subjects seeing a balanced range of harms in both the intentional and unintentional conditions.

Using a between-groups design, we also varied the emotional content of the scenarios independently of actor blameworthiness and harm. One group read scenarios that included gruesome, highly graphic language designed to boost emotional responses to the harm (graphic language (GL) group), whereas another group read scenarios that described the harm using just-the-facts language that avoided emotional content (plain language (PL) group) (Online Methods and **Supplementary Scenarios**). Crucially, the only difference between scenarios presented to the PL and GL groups was the language used to describe the harm, yet that language difference was sufficient to provoke stronger affective responses of disgust, contempt and sadness in the latter group (**Supplementary Fig. 1**). By manipulating only the emotional tenor of the language used to describe the harm (but not the harm itself), we are able not only to isolate the mechanisms by which affective heuristics can shape punishment severity, but also to determine how mental state information can overcome those influences.

RESULTS

Behavioral results

For all scenarios, participants were asked to assign the amount of punishment that they believed the scenario protagonist deserved for

[1]Center for Depression, Anxiety and Stress Research, McLean Hospital, Harvard Medical School, Belmont, Massachusetts, USA. [2]Department of Psychology, Emory University, Atlanta, Georgia, USA. [3]Department of Psychology, Harvard University, Cambridge, Massachusetts, USA. [4]University of Pennsylvania Law School, Philadelphia, Pennsylvania, USA. [5]Division of Social Sciences, Yale-NUS, Singapore. [6]Vanderbilt Law School, Nashville, Tennessee, USA. [7]Department of Biological Sciences, Vanderbilt University, Nashville, Tennessee, USA. [8]MacArthur Foundation Research Network on Law and Neuroscience, Vanderbilt University, Nashville, Tennessee, USA. [9]Department of Psychology, and Center for Integrative and Cognitive Neurosciences, Vanderbilt University, Nashville, Tennessee, USA. [10]These authors contributed equally to this work. Correspondence should be addressed to M.T.T. (mtreadway@emory.edu) or R.M. (rene.marois@vanderbilt.edu).

Received 21 May; accepted 10 July; published online 3 August 2014; doi:10.1038/nn.3781

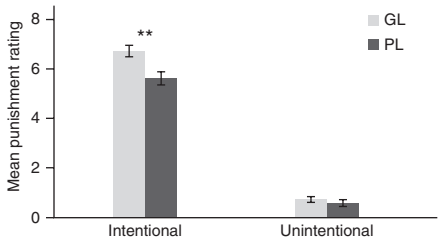

Figure 1 Effects of blameworthiness and graphic language on punishment ratings. There was a blameworthiness-by-language interaction, such that punishment ratings during intentional scenarios were significantly higher for the GL group as compared with the PL group, but there was no difference between these two groups for unintentional scenarios. Error bars represent s.e.m. $^{**}P < 0.01$.

his actions using a scale of 0 (no punishment) to 9 (harshest possible punishment)[15]. Across both the GL and PL groups, participants punished more (that is, assigned higher punishment ratings) during intentional scenarios than during unintentional scenarios ($F_{1,28} = 1367.69$, $P < 0.001$; **Fig. 1**). The magnitude of harm severity (for example, actions resulting in death versus loss of property) also predicted higher punishment ratings ($F_{1,84} = 215.44$, $P < 0.001$). Consistent with our primary hypothesis, participants in the GL group who were exposed to graphic descriptions of the harm punished more harshly than those in the PL group ($F_{1,28} = 9.50$, $P = 0.005$). Critically, however, a blameworthiness-by-language interaction was also observed ($F_{1,28} = 13.04$, $P = 0.001$), such that the GL group had higher punishment ratings for intentional scenarios ($t_{28} = 3.25$, $P = 0.003$), but not for unintentional scenarios ($t_{28} = 0.85$, $P = 0.40$) (**Fig. 1**). This blameworthiness-by-language interaction was not moderated by different levels of harm severity ($F_{1,28} = 0.58$, $P = 0.24$), and there were no differences in reaction time across groups or between conditions (all P values > 0.14; **Supplementary Fig. 2**). Taken together, these behavioral results support the hypothesis that the influence of emotional language on punishment severity is contingent on the perceived mental state of the actor.

fMRI results: BOLD amplitude

The GL contrast (GL versus PL groups averaged across blameworthiness conditions) did not yield any regional activations or deactivations after correcting for multiple comparisons. This is unsurprising considering that the behavioral effect of GL was largely confined to the intentional scenarios (see above and **Fig. 1**). Instead, based on our behavioral findings, the effect of GL on punishment-related brain activity should be strongly modulated by whether the protagonist intended the harms that he caused. Supporting this hypothesis, the interaction contrast of [(GL intentional > PL intentional) > (GL unintentional > PL unintentional)] revealed significant activity differences in bilateral amygdala (whole-brain cluster-corrected $P < 0.05$), which extended to anterior hippocampus and surrounding medial temporal cortex (Talairach coordinates, left: $x = -31$, $y = -8$, $z = -15$; right: $x = 23$, $y = -2$, $z = -18$; **Fig. 2**); other regions identified by this contrast included the left fusiform gyrus and bilateral superior temporal gyrus (**Supplementary Table 1**). Follow-up contrasts confirmed that the shape of the interaction mirrored that of the behavioral data. Specifically, the contrast of GL > PL for intentional scenarios revealed amygdala differences between the two groups ($P_{(\text{cluster-corrected})} < 0.05$), although these results only reached significance for the left amygdala, whereas there were no differences between the GL and PL groups in either left or right amygdala for unintentional trials.

The marked similarities between the pattern of activation in the left amygdala and the behavioral data suggest that this structure may be an important neural substrate underlying the blameworthiness-by-language interaction that we observed in the punishment ratings. If these changes in amygdala activity do affect punishment outcome, then this should be reflected in the pattern of connectivity between amygdala and key nodes of the punishment decision-making circuits. To test this hypothesis, we first sought to isolate areas involved in punishment decision-making, as these would be the necessary targets of any possible amygdalar influence. A key candidate region for punishment decision-making is the dorso-lateral prefrontal cortex (dlPFC, BA 9/46), as this area has been shown to be involved in both second-party and third-party punishment decisions and norm enforcement[15,17–19]. To identify such punishment decision-making regions, we used an analytical strategy that we have used previously[15], which consisted of contrasting activity between scenario conditions of high and low blameworthiness (**Fig. 3**). Here, using a contrast of intentional > unintentional, we found that intentional scenarios yielded greater activation in bilateral dlPFC (BA 9/46) (left: $x = -43$, $y = 34$, $z = 24$; right: $x = 41$, $y = 34$, $z = 27$; $P_{(\text{cluster-corrected})} < 0.05$; **Fig. 3a** and **Supplementary Table 2**). We confirmed that this region was involved in the decision to punish rather than reflecting the categorical differences in scenario type by contrasting punished versus unpunished trials during the unintentional condition only, as this contrast yielded the same activation pattern (**Fig. 3b** and **Supplementary Table 3**). Although the dlPFC showed a main effect of intentional versus unintentional scenarios, there was no effect of graphic language on the amplitude of the BOLD response. This result is not only consistent with prior findings[15], it is also consistent with the hypothesis that the role of these prefrontal areas, especially dlPFC, is to integrate multiple streams of information encoded by distinct nodes in the punishment decision-making network to select an appropriate response (integration-and-selection hypothesis[11]; J.W.B., J.W.M., M.T.T., K.J., D.H. Zald, O.D.J. & R.M., unpublished data), a process that needs not be reflected by monotonic amplitude differences as a function of punishment outcome[11].

fMRI results: Granger causality analyses

Our BOLD amplitude results replicate those of prior studies, implicate the dlPFC as a key node in a punishment-decision making network and suggest a plausible target by which GL-driven amygdala activity may influence punishment decisions. If this hypothesis is true, then

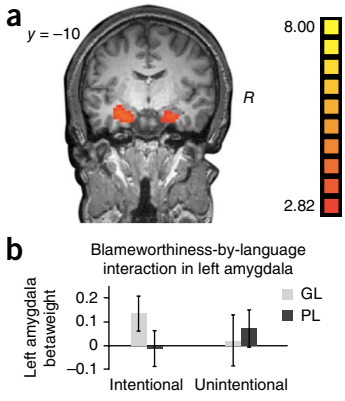

Figure 2 Amygdala activity mediates blameworthiness-by-language interaction during punishment decision-making. (**a**) SPM displaying the bilateral amygdala BOLD amplitude activation (rendered on a single-subject T1-weighted image), identified by the interaction contrast of [(GL intentional > PL intentional) > (GL unintentional > PL unintentional)], thresholded at $P < 0.05$ (cluster-corrected). (**b**) BOLD amplitude betaweights of left amygdala activity. Error bars represent s.e.m.

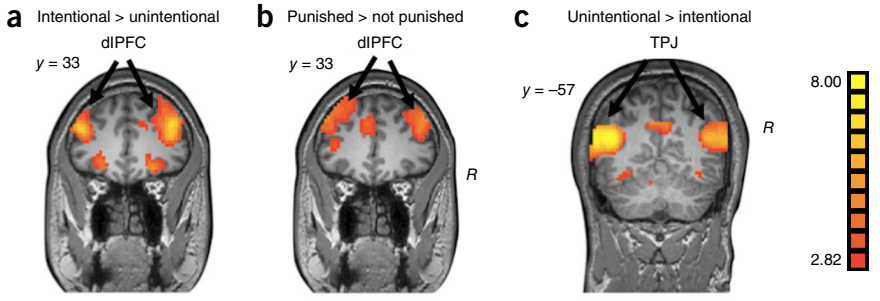

Figure 3 BOLD amplitude SPMs displaying dACC and dlPFC areas engaged in punishment decision-making (rendered on a single-subject T1-weighted image). (**a**,**b**) These regions were present when using either a main-effect contrast of intentional > unintentional scenarios (**a**) or a contrast of punished versus non-punished trials in the unintentional condition only (**b**). (**c**) Main-effect contrast of unintentional > intentional scenarios showing TPJ activation. Contrast averages across subjects from both GL and PL groups are shown. Map is displayed at a whole-brain corrected threshold of $P < 0.05$ (cluster corrected), rendered on a single-subject T1-weighted image.

we should be able to detect language group- and condition-specific changes in functional connectivity between the amygdala and this prefrontal area. To test this hypothesis, we used Granger causality mapping (GCM) analysis to examine condition differences in amygdala connectivity with the prefrontal cortex. GCM determines whether signals from a seed region serve as a better predictor of activity in a target region than past activity of the target region. As such, it can provide directional inferences about observed connectivity relationships (see Online Methods for further discussion of GCM analysis), although only in relation to the strength of the autoregressive model[21]. Because our pattern of behavioral data required the presence of both a blameworthiness-by-language interaction and a group difference for intentional scenarios alone, we focused our connectivity analyses on the left amygdala, as this was the only structure exhibiting this pattern.

Using the left amygdala as a seed, a blameworthiness-by-language interaction contrast of GCM connectivity revealed a cluster of left dlPFC (BA 9/46; $x = -38$, $y = 23$, $z = 19$) that received stronger input from the left amygdala ($P_{(\text{cluster-corrected})} < 0.05$; **Fig. 4a**). To confirm that the shape of this interaction was the same as our behavioral data, we performed follow-up group contrasts (GL > PL) for each condition separately. These analyses revealed that, relative to the PL group, the GL group showed stronger connectivity from left amygdala to left dlPFC (BA 9/46; $x = -33$, $y = 20$, $z = 20$) for intentional scenarios ($P_{(\text{cluster-corrected})} < 0.05$), but not for unintentional scenarios (**Fig. 4**). A conjunction analysis confirmed that the dlPFC area identified by the BOLD amplitude punishment decision-making analysis (intentional > unintentional) and the left dlPFC isolated by the amygdala-seed GCM (blameworthiness-by-language interaction) targeted the same region (**Supplementary Fig. 3a**).

Although these results provide a neural mechanism by which graphic language can produce harsher punishments for intentional acts, they do not reveal how such emotional influences are gated out when the harm was unintended. We hypothesized that, under such conditions, amygdala activity would be suppressed by regulatory regions that receive information about the actor's mental state. To test this hypothesis, we first examined whether any regions showed greater top-down signaling to the amygdala during a contrast of unintentional > intentional trials (target-to-seed connectivity). This contrast revealed several medial prefrontal and dorsomedial areas that exhibited stronger connectivity down to the amygdala in a contrast of unintentional > intentional trials across both language groups ($P_{(\text{cluster-corrected})} < 0.05$; **Supplementary Table 4**). Of the areas isolated by this contrast, the dorsal anterior cingulate cortex (dACC; peak voxel: $x = -10$, $y = 16$, $z = 30$; **Fig. 5**) is particularly relevant, as there is substantial evidence that this region exerts regulatory control over affect-encoding regions, including amygdala[22–24]. Indeed, the dACC cluster identified by our analysis was directly adjacent to the peak dACC coordinates isolated by a recent meta-analysis of emotion regulation studies[25]. There was no language group difference in this top-down dACC connectivity, which is consistent with both the lack of group differences in amygdala activity during the unintentional condition and the expectation that the downregulation of affective responses during unintentional scenarios should occur irrespective of the harm's description.

If it is true that the dACC is involved in gating harm-based affective arousal signals as a function of actor intent, then it should receive input from brain regions that have been implicated in inferring the mental states of others and in judgments of moral blameworthiness, such as the temporo-parietal junction (TPJ)[13–15]. Using the dACC as a seed in a second whole-brain GCM analysis, we found that, in a contrast of unintentional > intentional scenarios, the dACC received greater input (target-to-source connectivity) from left TPJ ($x = -47$, $y = -64$, $z = 4$; $P_{(\text{cluster-corrected})} < 0.05$). No other areas exhibited greater connectivity to dACC for this contrast. The identification of the TPJ as a source of dACC input in unintentional scenarios is consistent with the idea that this brain region's engagement in mental state inferences is enhanced when such inferences are more demanding, as when the scenario describes an incongruity between a mental state and a resulting harm[15]. In support of this assertion, we observed greater TPJ activation in unintentional scenarios relative to intentional scenarios ($P_{(\text{cluster-corrected})} < 0.05$; **Fig. 3c** and **Supplementary Table 2**). A follow-up conjunction analysis confirmed that the area of TPJ identified by our BOLD amplitude contrast of unintentional > intentional was the same as the area isolated by the dACC-seed GCM (**Supplementary Fig. 3b**).

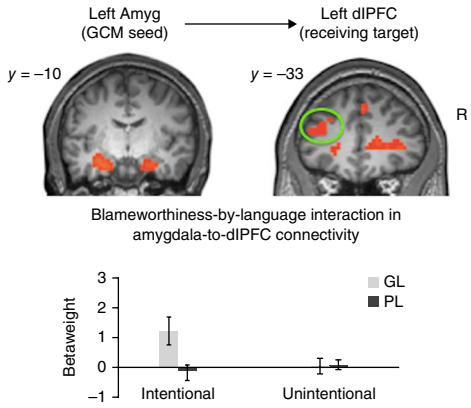

Figure 4 GCM of prefrontal regions influenced by amygdala seed. (**a**) GCM-based SPM showing region of left dlPFC identified by an interaction contrast [(GL intentional > PL intentional) > (GL unintentional > PL unintentional)] using the left amygdala as a seed. (**b**) Bar graph of GCM betaweights in left dlPFC. Positive values for GCM betaweights indicate connectivity from amygdala-to-target, whereas negative values indicate connectivity from target-to-amygdala. Scaling range of GCM betaweights are reported following application of a scaling factor of 100. Error bars represent s.e.m.

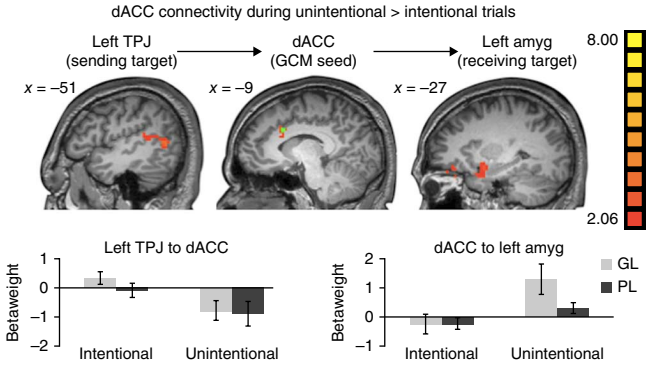

Figure 5 GCM-based connectivity identified by a contrast of unintentional > intentional scenarios based on dACC seed region. SPMs displaying TPJ-to-dACC connectivity (left) and dACC-to-amgydala connectivity (right) are shown. Center SPM shows the dACC cluster used to define the GCM seed region (green). Maps are shown at a whole-brain corrected threshold of $P < 0.05$ (cluster-corrected). Bar-whisker plots showing GCM beta weights averaged across GL and PL groups. For left TPJ, negative values for GCM betaweights indicate connectivity from target to dACC. For left amygdala, positive betaweight values indicate connectivity from dACC to the target. Scaling range of GCM betaweights are reported following application of a scaling factor of 100.

Taken together, these connectivity results complement our BOLD amplitude analyses in two important ways. First, they demonstrate that the increased amygdala activity observed in the GL group during intentional scenarios is associated with greater bottom-up amygdala-to-prefrontal connectivity, consistent with a heightened contribution of the amygdala to punishment decision-making in response to emotionally salient language. Second, they show that, regardless of the language used, there was greater top-down connectivity from the dACC to the amygdala during unintentional trials relative to intentional trials, suggesting a key role for this region in gating amygdala activity according to the perceived mental state. Consistent with this function, we observed that, during unintentional scenarios, the dACC received greater input from the TPJ, a region engaged by mentalizing. As such, this network reveals how knowledge of the actor's mental state may silence amygdala responses to the emotional content of unintended harms.

DISCUSSION

Although it is widely recognized that emotional heuristics are vital for human decision-making[26,27], the manner in which affective responses to harm and blameworthiness interact in punishment decisions had remained unaddressed. We found that information related to actor mental state regulates the influence of emotionally salient language on subsequent punishment and that this effect is mediated by a cortico-parieto-amygdalar circuit. Taken together, these findings reveal the functional neuro-architecture underlying the gating of affective-signals by intent during punishment decision-making.

A wealth of data has suggested that emotional content can alter judgments across a wide range of contexts, including moral blameworthiness[2] and legal punishment decision-making[3,28,29]. 'Affect-as-information' models posit that individuals use their emotional responses to help inform the judgment they are prompted to make[30–32]. However, these emotions are known to interact with evaluations of mental state in a variety of ways[33]. For instance, harms are experienced as more painful when they are perceived to be intentional[34], and unintended harms may go entirely unpunished, regardless of their magnitude[4,8,13,14]. By isolating one source of affect, graphic language descriptions of harm, in our design, we were able to demonstrate that manipulating emotional content was sufficient to increase punishment severity, but that its effect was contingent on perceptions of mental state.

This behavioral effect was reflected in brain activity. The amygdala showed a similar blameworthiness-by-language interaction, with greater activation in the GL group during intentional scenarios. Taken in isolation, however, this effect is insufficient to support a role for the amygdala in punishment decision-making, given evidence that final determination of punishment depends on dlPFC[15,17–20], a region that was strongly engaged by the decision to punish in the current study and that is generally involved in integration and selection among multiple sources of information[35,36]. What is required is evidence that amygdala activity affects the dlPFC during punishment decision-making. This was first provided by Granger causality analysis revealing a blameworthiness-by-language interaction such that greater connectivity from the left amygdala to left dlPFC was present in the GL group relative to the PL group, but only during intentional scenarios. It is important to note that direct anatomical connections between the amygdala and the dlPFC are relatively sparse[37,38]. However, the amygdala's reciprocal connections with multiple subdivisions of anterior cingulate and ventromedial PFC may produce downstream effects on dlPFC activity[39], which could explain frequent observations of functional coupling between these regions in human imaging studies[40,41].

Our study reveals that it is a different prefrontal region, however, the dACC, that has the most pivotal role in regulating affective brain regions as a function of mental state. During unintentional scenarios, the dACC exhibited top-down connectivity to amygdala, a result that may account for the abolishment of amygdala responses in this condition, as well as the lack of any language group differences in punishment. Furthermore, a critical input to dACC during unintentional scenarios was the TPJ, a region that encodes intentions in the context of harmful actions[13,14]. Emerging from these connectivity analyses is a crucial role for dACC in gating affective signals in response to harms that were caused in the absence of a blameworthy mental state. Functional imaging studies of the brain mechanisms for the cognitive control of emotion further implicate the dACC[42–44], especially when regulatory strategies rely on higher order reasoning, such as inferring mental states[22,42,44].

It is notable that the amplitude and connectivity results in the amygdala were primarily lateralized to the left hemisphere. Although this lateralization was not explicitly hypothesized, it is not altogether surprising. Past meta-analyses of amygdala activation suggest that the left amygdala responds more strongly to affective stimuli[45]. This left-lateralization is especially prominent in situations in which stimuli are negatively valenced[46] and language based[47], both of which apply to the current procedure.

Although our study elucidates the mechanisms by which inferred mental state can trump emotional responses to harm, our results should not be taken to suggest that harm is the only source of negative emotions or that the relationships between affect and mental state are unidirectional. Perceiving a person to have malicious intent can also be a substantial source of retributive emotions[2,9,48], and affect-induction procedures have been used to show that higher levels of negative emotions can result in stronger attributions of moral wrongfulness[49]. For this reason, our blameworthiness manipulation was designed to be categorically unambiguous, with the harm caused by the protagonist always being described as either completely intended or utterly unintended. In addition, harm information was held constant across both GL and PL groups. By effectively controlling for these other sources of

affect, our design allowed us to isolate the brain mechanisms by which inferences about mental state can shunt emotional urges to punish, a core principle in determining deserved punishment.

That being said, these findings may very well generalize to other sources of affect. Indeed, a blameworthiness-by-harm interaction analysis revealed the same left amygdala area highlighted by the blameworthiness × GL interaction, further suggesting the need for amygdala downregulation in both groups during unintentional scenarios (**Supplementary Fig. 4**). The blameworthiness-by-harm interaction was equally strong for both the GL and PL groups, suggesting that, regardless of the language used, the amygdala is broadly involved in the encoding of harm in a blameworthiness-dependent manner[15]. It is therefore tempting to speculate that the amygdala may ultimately be involved in encoding one's sense of moral outrage, a consolidated heuristic of our anger to wrongdoing[50]. Ultimately, however, the crux of the present study is not so much about the specific computation carried out by the amygdala in punishment decision-making as it is about how this brain region, along with those involved in affect regulation, theory of mind and cognitive selection form a regulatory network that governs the interacting influences of inferred mental state and emotional responses in our decisions to punish.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

The authors would also like to acknowledge helpful comments made by N.A. Farahany on the development of the scenario stimuli. Preparation of this article was supported by an award to the Vanderbilt University Central Discovery Grant Program to R.M. and O.D.J., as well as contributions from the John D. and Catherine T. MacArthur Foundation. Its contents reflect the views of the authors and do not necessarily represent the official views of either the John D. and Catherine T. MacArthur Foundation or The MacArthur Foundation Research Network on Law and Neuroscience. The authors also gratefully acknowledge support from the Center for Integrative and Cognitive Neuroscience at Vanderbilt University as well as support by the National Center for Research Resources, Grant UL1 RR024975-01, which is now at the National Center for Advancing Translational Sciences, Grant 2 UL1 TR000445-06.

AUTHOR CONTRIBUTIONS

M.T.T., J.W.B. and R.M. designed the study. M.T.T., J.W.M., K.J., J.W.B., O.D.J. and R.M. developed the scenario stimuli. M.T.T., J.W.M., K.J., J.W.B., M.R.G. and R.M. collected and analyzed the data with the aid of critical tools provided by C.L.A. M.T.T., J.W.B. and R.M. drafted the paper. J.W.M. and O.D.J. provided critical comments.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Greene, J. & Haidt, J. How (and where) does moral judgment work? *Trends Cogn. Sci.* **6**, 517–523 (2002).
2. Haidt, J. The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychol. Rev.* **108**, 814 (2001).
3. Bright, D.A. & Goodman-Delahunty, J. Gruesome evidence and emotion: anger, blame and jury decision-making. *Law Hum. Behav.* **30**, 183–202 (2006).
4. Darley, J.M. Morality in the law: the psychological foundations of Citizens' desires to punish transgressions. *Ann. Rev. Law Soc. Sci.* **5**, 1–23 (2009).
5. Goldberg, J.H., Lerner, J.S. & Tetlock, P.E. Rage and reason: the psychology of the intuitive prosecutor. *Eur. J. Soc. Psychol.* **29**, 781–795 (1999).
6. Shen, F., Hoffman, M., Jones, O., Greene, J. & Marois, R. Sorting guilty minds. *N. Y. Univ. Law Rev.* **86**, 1–55 (2011).
7. Cushman, F. Crime and punishment: distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition* **108**, 353–380 (2008).
8. Alter, A.L., Kernochan, J. & Darley, J.M. Transgression wrongfulness outweighs its harmfulness as a determinant of sentence severity. *Law Hum. Behav.* **31**, 319–335 (2007).
9. Darley, J.M. & Pittman, T.S. The psychology of compensatory and retributive justice. *Pers. Soc. Psychol. Rev.* **7**, 324–336 (2003).
10. Shenhav, A. & Greene, J.D. Moral judgments recruit domain-general valuation mechanisms to integrate representations of probability and magnitude. *Neuron* **67**, 667–677 (2010).
11. Buckholtz, J.W. & Marois, R. The roots of modern justice: cognitive and neural foundations of social norms and their enforcement. *Nat. Neurosci.* **15**, 655–661 (2012).
12. Feigenson, N. & Park, J. Emotions and attributions of legal responsibility and blame: a research review. *Law Hum. Behav.* **30**, 143–161 (2006).
13. Young, L., Camprodon, J.A., Hauser, M., Pascual-Leone, A. & Saxe, R. Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. *Proc. Natl. Acad. Sci. USA* **107**, 6753–6758 (2010).
14. Young, L., Cushman, F., Hauser, M. & Saxe, R. The neural basis of the interaction between theory of mind and moral judgment. *Proc. Natl. Acad. Sci. USA* **104**, 8235–8240 (2007).
15. Buckholtz, J.W. *et al.* The neural correlates of third-party punishment. *Neuron* **60**, 930–940 (2008).
16. Heekeren, H.R. *et al.* Influence of bodily harm on neural correlates of semantic and moral decision-making. *Neuroimage* **24**, 887–897 (2005).
17. Sanfey, A.G., Rilling, J.K., Aronson, J.A., Nystrom, L.E. & Cohen, J.D. The neural basis of economic decision-making in the Ultimatum Game. *Science* **300**, 1755–1758 (2003).
18. Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V. & Fehr, E. Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science* **314**, 829–832 (2006).
19. Spitzer, M., Fischbacher, U., Herrnberger, B., Gron, G. & Fehr, E. The neural signature of social norm compliance. *Neuron* **56**, 185–196 (2007).
20. Greene, J.D., Nystrom, L.E., Engell, A.D., Darley, J.M. & Cohen, J.D. The neural bases of cognitive conflict and control in moral judgment. *Neuron* **44**, 389–400 (2004).
21. Goebel, R., Roebroeck, A., Kim, D.S. & Formisano, E. Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping. *Magn. Reson. Imaging* **21**, 1251–1261 (2003).
22. Ochsner, K.N. *et al.* For better or for worse: neural systems supporting the cognitive down- and up-regulation of negative emotion. *Neuroimage* **23**, 483–499 (2004).
23. Urry, H.L. *et al.* Amygdala and ventromedial prefrontal cortex are inversely coupled during regulation of negative affect and predict the diurnal pattern of cortisol secretion among older adults. *J. Neurosci.* **26**, 4415–4425 (2006).
24. Banks, S.J., Eddy, K.T., Angstadt, M., Nathan, P.J. & Phan, K.L. Amygdala-frontal connectivity during emotion regulation. *Soc. Cogn. Affect. Neurosci.* **2**, 303–312 (2007).
25. Buhle, J.T. *et al.* Cognitive reappraisal of emotion: a meta-analysis of human neuroimaging studies. *Cereb. Cortex* published online doi:10.1093/cercor/bht154 (13 June 2012).
26. Loewenstein, G. & Lerner, J.S. The role of affect in decision making. In *Handbook of Affective Science* (eds. Davdison, R., Goldsmith, H. & Scherer, K.) 619–642 (2003).
27. Damasio, A. *Descartes' Error: Emotion, Reason and the Human Brain* (Penguin, 2005).
28. Bright, D.A. & Goodman-Delahunty, J. The influence of verbal gruesome evidence on mock juror verdicts. *Psychiatry Psychol. Law* **11**, 154–166 (2004).
29. Douglas, K.S., Lyon, D.R. & Ogloff, J.R. The impact of graphic photographic evidence on mock jurors' decisions in a murder trial: probative or prejudicial? *Law Hum. Behav.* **21**, 485–501 (1997).
30. Clore, G.L. & Storebeck, J. Affect as information in social judgments and behaviors. in *Hearts and Minds: Affective Influences on Social Thinking and Behavior* (ed. J.P. Forgas) (Psychological Press, 2006).
31. Schwarz, N. Feelings as information: Informational and motivational functions of affective states. in *Handbook of Motivation and Cognition* (eds. E.T. Higgins & R. Sorrentino) (Guilford Press, New York, 1990).
32. Forgas, J.P. Mood and judgment: the affect infusion model (AIM). *Psychol. Bull.* **117**, 39 (1995).
33. Weiner, B. *Judgements of Responsibility: a Foundation For a Theory of Social Conduct* (Guilford Press, 1995).
34. Gray, K. & Wegner, D.M. The sting of intentional pain. *Psychol. Sci.* **19**, 1260–1262 (2008).
35. Banich, M.T. *et al.* Cognitive control mechanisms, emotion and memory: a neural perspective with implications for psychopathology. *Neurosci. Biobehav. Rev.* **33**, 613–630 (2009).
36. Miller, E.K. & Cohen, J.D. An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* **24**, 167–202 (2001).
37. Stefanacci, L. & Amaral, D.G. Topographic organization of cortical inputs to the lateral nucleus of the macaque monkey amygdala: a retrograde tracing study. *J. Comp. Neurol.* **421**, 52–79 (2000).

38. Ghashghaei, H.T. & Barbas, H. Pathways for emotion: interactions of prefrontal and anterior temporal pathways in the amygdala of the rhesus monkey. *Neuroscience* **115**, 1261–1279 (2002).
39. Medalla, M. & Barbas, H. Synapses with inhibitory neurons differentiate anterior cingulate from dorsolateral prefrontal pathways associated with cognitive control. *Neuron* **61**, 609–620 (2009).
40. Anticevic, A., Repovs, G. & Barch, D.M. Emotion effects on attention, amygdala activation, and functional connectivity in schizophrenia. *Schizophr. Bull.* **38**, 967–980 (2012).
41. Siegle, G.J., Thompson, W., Carter, C.S., Steinhauer, S.R. & Thase, M.E. Increased amygdala and decreased dorsolateral prefrontal BOLD responses in unipolar depression: related and independent features. *Biol. Psychiatry* **61**, 198–209 (2007).
42. Diekhof, E.K., Geier, K., Falkai, P. & Gruber, O. Fear is only as deep as the mind allows: a coordinate-based meta-analysis of neuroimaging studies on the regulation of negative affect. *Neuroimage* **58**, 275–285 (2011).
43. Hartley, C.A. & Phelps, E.A. Changing fear: the neurocircuitry of emotion regulation. *Neuropsychopharmacology* **35**, 136–146 (2010).
44. Ochsner, K.N. & Gross, J.J. The cognitive control of emotion. *Trends Cogn. Sci.* **9**, 242–249 (2005).
45. Baas, D., Aleman, A. & Kahn, R.S. Lateralization of amygdala activation: a systematic review of functional neuroimaging studies. *Brain Res. Brain Res. Rev.* **45**, 96–103 (2004).
46. Fusar-Poli, P. *et al.* Laterality effect on emotional faces processing: ALE meta-analysis of evidence. *Neurosci. Lett.* **452**, 262–267 (2009).
47. Costafreda, S.G., Brammer, M.J., David, A.S. & Fu, C.H. Predictors of amygdala activation during the processing of emotional stimuli: a meta-analysis of 385 PET and fMRI studies. *Brain Res. Rev.* **58**, 57–70 (2008).
48. Schnall, S., Haidt, J., Clore, G.L. & Jordan, A.H. Disgust as embodied moral judgment. *Pers. Soc. Psychol. Bull.* **34**, 1096–1109 (2008).
49. Wheatley, T. & Haidt, J. Hypnotic disgust makes moral judgments more severe. *Psychol. Sci.* **16**, 780–784 (2005).
50. Carlsmith, K.M., Darley, J.M. & Robinson, P.H. Why do we punish? Deterrence and just deserts as motives for punishment. *J. Pers. Soc. Psychol.* **83**, 284–299 (2002).

ONLINE METHODS

Participants. A total of 37 healthy community volunteers were recruited to participate in this study. All participants provided written informed consent. All study procedures were approved by the Vanderbilt University Institutional Review Board and all data were collected at Vanderbilt. All subjects were required to be right-handed and possess normal or corrected vision. Subjects were excluded if they had any condition that would interfere with an MRI scan (for example, claustrophobia, cochlear implant, cardiac pacemaker), and if they reported any substantial trauma related to a crime that might make reading the scenario stimuli too upsetting. If subjects were determined to be eligible, they were assigned to either the PL or GL group in an alternating fashion.

Target sample size was determined to be 15 per group based on effects sizes from our prior study[15] and effect sizes during pilot data collection for the current scenario stimuli. Data from 7 subjects could not be used due to technical problems with scanning acquisition or excessive head motion (>3 mm), resulting in a final sample of 30 subjects (15 per group). Within these 30 subjects, data from 3 subjects was based on 8 of the 9 runs due to either excess motion or equipment failure for one of the runs. For these 30 subjects, 20 were male. Average age was 22.8 years, with a range of 18 to 30. There were no significant differences in age ($t_{28} = -0.65$, $P = 0.52$) or gender ($\chi^2_{(1)}$ 2.4, $P = 0.25$) between the two groups.

Experimental stimuli. Experimental stimuli were text-based scenarios. A list of all scenarios used in the study is provided in the **Supplementary Scenarios**. Each scenario was based on one of the 68 scenario 'stems' that depicted a protagonist (John) whose actions brought about harm to one other person (Steve or Mary). For each stem, four scenario variations were created to reflect two levels of each of the two primary variables of interest: criminal mental state and language description. Specifically, the four variations for each scenario stem were: graphic language/blameworthy harm (GL-intentional), graphic language/unintentional harm (GL-unintentional), plain language/blameworthy harm (PL-intentional) and plain language/unintentional harm (PL-unintentional). All other scenario content was identical across the four versions of a theme. Each subject saw a given scenario stem only once, with the assignment of a particular scenario stem to one of the four variations counterbalanced across subjects. To control for the possibility that graphic language might influence perceptions of mental state[3,5,28,29,51], all scenarios were designed to be categorically unambiguous with respect to John's mental state. In addition, the actual magnitude of harm was identical across the GL and PL conditions, although it was described in more graphic language in the GL condition. Below is an example of the differences between PL and GL descriptions of harm:

GL scenario: "John and Steve are avid mountain climbers and often climb together. John secretly wants to date Steve's girlfriend, and plans to kill Steve the next time the two of them go climbing together. Several weeks later, John and Steve go climbing. As they are rappelling down the face of a steep cliff, John takes a knife to Steve's rigging and severs Steve's support lines. Steve plummets to the rocks below. Nearly every bone in his body is broken upon impact. Steve's screams are muffled by thick, foamy blood flowing from his mouth as he bleeds to death."

PL scenario: "John and Steve are avid mountain climbers and often climb together. John secretly wants to date Steve's girlfriend, and plans to kill Steve the next time the two of them go climbing together. Several weeks later, John and Steve go climbing. As they are rappelling down the face of a steep cliff, John takes a knife to Steve's rigging and severs Steve's support lines. Steve falls one hundred feet to the ground below. Steve experiences significant bodily harm from the fall, and he dies from his injuries shortly after impact."

Scenarios also varied in terms of level of harm that befell the victim. The four harm categories included death, maiming, physical assault and property damage. All subjects saw a total of 68 scenarios, 34 intentional and 34 unintentional, with equal numbers of intentional and unintentional scenarios distributed equally among 6 murder scenarios, 16 maim scenarios, 6 assault and 6 property damage scenarios. Maim scenarios were used more frequently as they provided the easiest format in which to vary the graphic language used between the GL and PL conditions.

This experiment consisted of a 2 × 2 × 4 design, with blameworthiness (2) and harm (4) as within-subjects factors, and GL (2) as a between-subjects factor. Specifically, half the subjects ($n = 15$) saw only PL scenarios, whereas the other half ($n = 15$) saw only GL scenarios. We used a between-subjects design for this manipulation given concerns that subjects who view both GL and PL scenarios repeatedly might intuit the purpose of the experiment, leading to demand characteristics. Moreover, past research has shown the presence of 'carry-over' effects from one scenario to another, where emotions of anger evoked by one scenario may influence punishment on a subsequent scenario[52]. Such potential carry-over effects would be particularly detrimental to the language manipulation as the GL and PL conditions would tend to neutralize each other. We note that, although carry-over effects may still have occurred across intentional and unintentional scenarios, these were minimized by presenting scenarios in a pseudo-randomized order for each subject. All scenarios were equivalent in word count across all four versions (all P values > 0.3). In addition, conditions did not differ in regards to the temporal occurrence of the harm description in the scenarios, such that the amount of words that occurred before the description of the harm was matched across scenarios (all P values > 0.3). This additional control ensures against any imbalance in the timing of harm presentation across the conditions, which could confound interpretations of BOLD signals.

Experimental protocol. Subjects were told that the purpose of the experiment was to understand the brain mechanisms involved in assigning punishment, but were given no indication that there was a GL manipulation. Following debriefing, no subjects reported any awareness of the study's true hypotheses. For each scenario, subjects were asked to rate how much punishment they believed that John deserved for his actions. Punishment ratings were made using a 0–9 scale where 0 signified no punishment and 9 signified the most severe punishment that the subject endorsed. Subjects were not, however, explicitly instructed to decide how their scale corresponded to a specific punishment type.

Prior to scanning, all subjects viewed eight practice scenarios that spanned the whole range of harm, intent and language. Scenarios were presented during scanning using a visual display presented on an LCD panel and back-projected onto a screen positioned at the front of the magnet bore. Subjects were positioned supine in the scanner so as to be able to view the projector display using a mirror above their eyes. Manual responses were recorded using two five-button keypads (one for each hand, Rowland Institute of Science). Subjects were instructed to make a manual response as soon as they had arrived at a punishment decision, so as to ensure that neural activity around the time of response would reflect decision-making. After each manual-press, subjects viewed a fixation cue for a 12 s intertrial interval. Scanning was completed by a scanner technician and a research assistant. The research assistant was not blind to group assignment.

Statistical analysis: behavioral data. Behavioral data (punishment ratings and reaction time data) for intentional and unintentional scenarios presented to the PL and GL groups were analyzed using SPSS 21 (IBM). All statistical tests used are two-sided, unless otherwise specified. Distribution normality for each variable was assessed using standard metrics of skewness to ensure suitability of parametric statistical tests, with a cutoff of ±1.25. In addition, all data were inspected for outliers or unduly influential points ($z = \pm 3.5$). For assessment of the blameworthiness-by-emotion interaction, a mixed-design ANOVA was used. This interaction was then decomposed using follow-up t tests. Differences in variance across the two groups were assessed using either Mauchly's sphercity test (ANOVA) or a levene test (t tests). All reported results met assumptions of homoscedasticity for parametric tests.

fMRI data acquisition. All fMRI scans were acquired using a 3T Philips Achieva scanner at the Vanderbilt University Institute of Imaging Science. Stimulus presentation was synchronized to fMRI volume acquisition. Low- and high-resolution structural scans were first acquired using conventional parameters. Functional (T2* weighted) images were acquired using a gradient-echo echoplanar imaging (EPI) pulse sequence with the following parameters: TR = 2,000 ms, TE = 57 ms, flip angle = 90°, FOV = 240 × 240 mm, 128 × 128 matrix with 35 axial slices (2.5 mm, 0.5 mm gap) oriented at a 15° oblique angle to the AC-PC. This slice prescription was selected for optimization of BOLD signal in the amygdala, given that this structure was one of the primary regions of interest for this study. Each scanning session included 9 functional runs. The first 8 runs lasted 9 min, and contained 12 trials each. The last run contained only 4 trials, and lasted 3 min to ensure all participants viewed a balanced number of intentional and unintentional scenarios.

Statistical analysis: fMRI data. Image analysis was conducted using Brain Voyager QX 2.3 (Brain Innovation) in conjunction with custom Matlab software. All images were preprocessed using slice timing correction, three-dimensional motion correction, linear trend removal, and spatial smoothing with an 8 mm Gaussian kernel (full-width at half-maximum) as implemented through Brain Voyager software. Subjects' functional data were co-registered with their T1-weighted anatomical volumes and transformed into standardized Talairach space. To ensure that preprocessing steps rendered data suitable for GCM analysis, linear and slow-wave components were removed using a high-pass filter.

For each subject, design matrices for a fixed-effects general linear model (GLM) was conducted by convolving a canonical hemodynamic response function (double gamma, including a positive γ function and a smaller, negative γ function to reflect the BOLD undershoot) to the following set of regressors: a baseline that represented all signal acquired during the inter-trial interval (baseline), the onset of a new scenario (reading phase), and the time point 3 TRs (6s) before the subject's response plus the TR of the subject's response (decision-phase). The division of scenario presentation into a reading phase and a decision phase was modeled after studies using scenario-based presentation as means of exploring neural correlates of social decision-making (for example, see ref. 20). In keeping with other neuroimaging studies of decision-making[15,20,53], decision-related modulation of BOLD signal would be expected to correspond with the portion of the time course around the subject's response. After each scenario, subjects viewed a small white square on the screen for a jittered inter-trial interval that ranged between 10.8 and 14.8 s, allowing sufficient time for the hemodynamic response to return to baseline. All fixed-effects models included an AR(1) term to account for within-subject correlations. Beta weights for each fMRI run were transformed into *z* scores signifying the magnitude of deviation of the fMRI signal during either the reading-phase or decision-phase as compared to the average signal during the intertrial interval period.

Second-order random effects analyses were conducted by contrasting the beta-weights from each subject's fixed-effects analyses in a single GLM model generated using a 2 (language) by 2 (blameworthiness) by 2 (scenario phase) by 4 (harm) design matrix. All analyses focused on activation during the decision phase. Based on our a priori hypotheses, the primary contrast of interest was $[(GL_{intentional} - GL_{unintentional}) - (PL_{intentional} - PL_{unintentional})]$. To control for multiple comparisons, a Forman correction procedure[54] was implemented to determine appropriate voxel-height and spatial-extent procedures to maintain a whole-brain false-positive error rate of $P < 0.05$ for all random-effects GLMs and GCMs, as implemented by the ClusterThresh plugin to BrainVoyager. To determine the minimum cluster-threshold for each analysis, a MonteCarlo simulation was generated for each contrast. Only results surviving this threshold are reported.

After identifying all functionally defined ROIs, GLM betaweights were extracted and averaged across condition and group for display purposes. To avoid any circular or non-independent analyses[55], all inferential statistics reported were performed in neuroimaging space using whole-brain corrected statistical thresholds and no secondary inferential statistical tests were performed on data extracted from ROIs. Extracted ROI data presented in figures are non-independent and should not be used for effect-size estimates, but are included as a visual aid for the interpretation of results from statistical analyses performed in neuroimaging space[56].

Granger causality mapping. Granger causality is a multivariate autoregressive technique that can be used to test prediction models in time-series data[57], including fMRI data[58–60]. All GCM analyses were implemented using the GCM plug-in developed for Brainvoyager, for which detailed methods have been published previously[21,61] (http://support.brainvoyager.com/documents/Functional_Statistics/GCM/main.html). Briefly, GCM compares the time courses of a seed region $\boldsymbol{x}[n]$ and a target region $\boldsymbol{y}[n]$ to determine the incremental predictive power of including x in the prediction y against a vector autoregressive (AR) model to predict that $\boldsymbol{y}[n]$. The order of the AR model was specified to be 1 (refs. 61–63). To isolate connectivity associated with decision-making (as compared to intrinsic connectivity), a trimmed-time series analysis was used[58], which involved isolating TRs during the decision-phase of each trial for each condition. For the amygdala GCM analysis, the seed region used was a 6mm cube centered on the voxel-peak in the left amygdala identified by our GLM interaction contrast. For the dACC GCM analysis, the seed region was a 6mm cube centered on the voxel peak of the dACC as identified by the GL>PL contrast during intentional scenarios from the amygdala GCM map.

The results of GCM analysis include either positive values, where $\boldsymbol{x}[n]$ offers superior prediction of activity as compared to $\boldsymbol{y}[n]$ (seed-to-target connectivity), or negative values, where information about $\boldsymbol{y}[n]$ provides improved prediction of $\boldsymbol{x}[n]$ (target-to-seed connectivity). All GCM maps were interrogated using a whole-brain, cluster-corrected threshold. *P* values were obtained using a bootstrapping procedure with 5,000 simulations. As with our random effects analysis, correction for multiple comparisons in our GCM analyses was achieved using a Forman correction, as described above. In addition, it should be noted that some recent work has suggested that resting-state GCM analyses may reflect oxidation differences across cerebrovasculature, rather than regional activity[64]. To deal with this issue, it has been proposed that within-subject condition contrasts of GCM betaweights within the same region can provide an effective control against possible blood flow effects[61] (see also http://www.russpoldrack.org/2013/12/a-discussion-of-causal-inference-on.html for a discussion). Consequently, all GCM analyses presented herein are based on within-subject condition contrasts of GCM betaweights within the same region.

A **Supplementary Methods** Checklist is available.

51. Kassin, S.M. & Garfield, D.A. Blood and guts: general and trial specific effects of videotaped crime scenes on mock jurors. *J. Appl. Soc. Psychol.* **37**, 1877–1887 (1991).
52. Goldberg, J.H., Lerner, J.S. & Tetlock, P.E. Rage and reason: the psychology of the intuitive prosecutor. *Eur. J. Soc. Psychol.* **29**, 781–795 (1999).
53. Aron, A.R. & Poldrack, R.A. Cortical and subcortical contributions to Stop signal response inhibition: role of the subthalamic nucleus. *J. Neurosci.* **26**, 2424–2433 (2006).
54. Forman, S.D. *et al.* Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magn. Reson. Med.* **33**, 636–647 (1995).
55. Kriegeskorte, N., Simmons, W.K., Bellgowan, P.S. & Baker, C.I. Circular analysis in systems neuroscience: the dangers of double dipping. *Nat. Neurosci.* **12**, 535–540 (2009).
56. Poldrack, R.A. & Mumford, J.A. Independence in ROI analysis: where is the voodoo? *Soc. Cogn. Affect. Neurosci.* **4**, 208–213 (2009).
57. Granger, C.W. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**, 424–438 (1969).
58. Rogers, B.P., Morgan, V.L., Newton, A.T. & Gore, J.C. Assessing functional connectivity in the human brain by fMRI. *Magn. Reson. Imaging* **25**, 1347–1357 (2007).
59. Harrison, L., Penny, W.D. & Friston, K. Multivariate autoregressive modeling of fMRI time series. *Neuroimage* **19**, 1477–1491 (2003).
60. Kamiński, M., Ding, M., Truccolo, W.A. & Bressler, S.L. Evaluating causal relations in neural systems: granger causality, directed transfer function and statistical assessment of significance. *Biol. Cybern.* **85**, 145–157 (2001).
61. Roebroeck, A., Formisano, E. & Goebel, R. Mapping directed influence over the brain using Granger causality and fMRI. *Neuroimage* **25**, 230–242 (2005).
62. Wen, X., Yao, L., Liu, Y. & Ding, M. Causal interactions in attention networks predict behavioral performance. *J. Neurosci.* **32**, 1284–1292 (2012).
63. Hamilton, J.P., Chen, G., Thomason, M.E., Schwartz, M.E. & Gotlib, I.H. Investigating neural primacy in Major Depressive Disorder: multivariate Granger causality analysis of resting-state fMRI time-series data. *Mol. Psychiatry* **16**, 763–772 (2011).
64. Webb, J.T., Ferguson, M.A., Nielsen, J.A. & Anderson, J.S. BOLD granger causality reflects vascular anatomy. *PLoS One* **8**, e84279 (2013).

doi:10.1038/nn.3781