doi:10.1093/scan/nss069 SCAN (2012) 7, 743–751

Differential neural circuitry and self-interest in real *vs* hypothetical moral decisions

Oriel Feldman Hall,[1,2] Tim Dalgleish,[1] Russell Thompson,[1] Davy Evans,[1,2] Susanne Schweizer,[1,2] and Dean Mobbs[1]
[1]Medical Research Council, Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, UK and [2]Cambridge University, Cambridge CB2 1TP, UK

Classic social psychology studies demonstrate that people can behave in ways that contradict their intentions—especially within the moral domain. We measured brain activity while subjects decided between financial self-benefit (earning money) and preventing physical harm (applying an electric shock) to a confederate under both real and hypothetical conditions. We found a shared neural network associated with empathic concern for both types of decisions. However, hypothetical and real moral decisions also recruited distinct neural circuitry: hypothetical moral decisions mapped closely onto the imagination network, while real moral decisions elicited activity in the bilateral amygdala and anterior cingulate—areas essential for social and affective processes. Moreover, during real moral decision-making, distinct regions of the prefrontal cortex (PFC) determined whether subjects make selfish or pro-social moral choices. Together, these results reveal not only differential neural mechanisms for real and hypothetical moral decisions but also that the nature of real moral decisions can be predicted by dissociable networks within the PFC.

Keywords: real moral decision-making; fMRI; amygdala; TPJ; ACC

INTRODUCTION

Psychology has a long tradition demonstrating a fundamental difference between how people believe they will act and how they actually act in the real world (Milgram, 1963; Higgins, 1987). Recent research (Ajzen et al., 2004; Kang et al., 2011; Teper et al., 2011) has confirmed this intention–behavior discrepancy, revealing that people inaccurately predict their future actions because hypothetical decision-making requires mental simulations that are abbreviated, unrepresentative and decontextualized (Gilbert and Wilson, 2007). This 'hypothetical bias' effect (Kang et al., 2011) has routinely demonstrated that the influence of socio-emotional factors and tangible risk (Wilson et al., 2000) is relatively diluted in hypothetical decisions: not only do hypothetical moral probes lack the tension engendered by competing, real-world emotional choices but also they fail to elicit expectations of consequences—both of which are endemic to real moral reasoning (Krebs et al., 1997). In fact, research has shown that when real contextual pressures and their associated consequences come into play, people can behave in characteristically immoral ways (Baumgartner et al., 2009; Greene and Paxton, 2009). Although there is also important work examining the neural basis of the opposite behavioral finding—altruistic decision-making (Moll et al., 2006)—the neural networks underlying the conflicting motivation of maximizing self-gain at the expense of another are still poorly understood.

Studying the neural architecture of this form of moral tension is particularly compelling because monetary incentives to behave immorally are pervasive throughout society—people frequently cheat on their loved ones, steal from their employers or harm others for monetary gain. Moreover, we reasoned that any behavioral and neural disparities between real and hypothetical moral reasoning will likely have the sharpest focus when two fundamental proscriptions—do not harm others and do not over-benefit the self at the expense of others (Haidt, 2007)—are directly pitted against one another. In other words, we speculated that this prototypical moral conflict would provide an ideal test-bed to examine the behavioral and neural differences between intentions and actions.

Received 18 April 2012; Accepted 8 June 2012
Advance Access publication 18 June 2012

Correspondence should be addressed to Oriel FeldmanHall, MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, UK. E-mail: Oriel.FeldmanHall@mrc-cbu.cam.ac.uk

Accordingly, we used a 'your pain, my gain' (PvG) laboratory task (Feldmanhall et al., 2012) to operationalize this core choice between personal advantage and another's welfare: subjects were probed about their willingness to receive money (up to £200) by physically harming (via electric stimulations) another subject (Figure 1A). The juxtaposition of these two conflicting motivations requires balancing selfish needs against the notion of 'doing the right thing' (Blair, 2007). We carried out a functional magnetic resonance imaging (fMRI) experiment using the PvG task to first explore if real moral behavior mirrors hypothetical intention, and second, to examine if these two classes of behavior are subserved by the same neural architecture. We hypothesized that people would imagine doing one thing, but when faced with real monetary incentive, do another—and that this behavioral difference would be reflected at the neurobiological level with differential patterns of activity.

MATERIALS AND METHODS

Subjects

Fourteen healthy subjects took part in this study: six males; mean age and s.d. 25.9 ± 4.6, completed a Real PvG, Imagine PvG and a Non-Moral control task in a within-subject design while undergoing fMRI. Four additional subjects were excluded from analyzes due to expressing doubts about the veracity of the Real PvG task on a post-scan questionnaire and during debriefing. Two additional subjects were not included because of errors in acquiring scanning images. Subjects were compensated for their time and travel and allowed to keep any earnings accumulated during the task. All subjects were right-handed, had normal or corrected vision and were screened to ensure no history of psychiatric or neurological problems. All subjects gave informed consent, and the study was approved by the University of Cambridge, Department of Psychology Research Ethics Committee.

Experimental tasks

Real pain vs gain task (Real PvG)

In the Real PvG subjects (Deciders) were given £20 and asked how much of their money they were willing to give up to prevent a series of painful electric stimulations from reaching the wrist of the second subject (the Receiver—a confederate). The more money the Decider

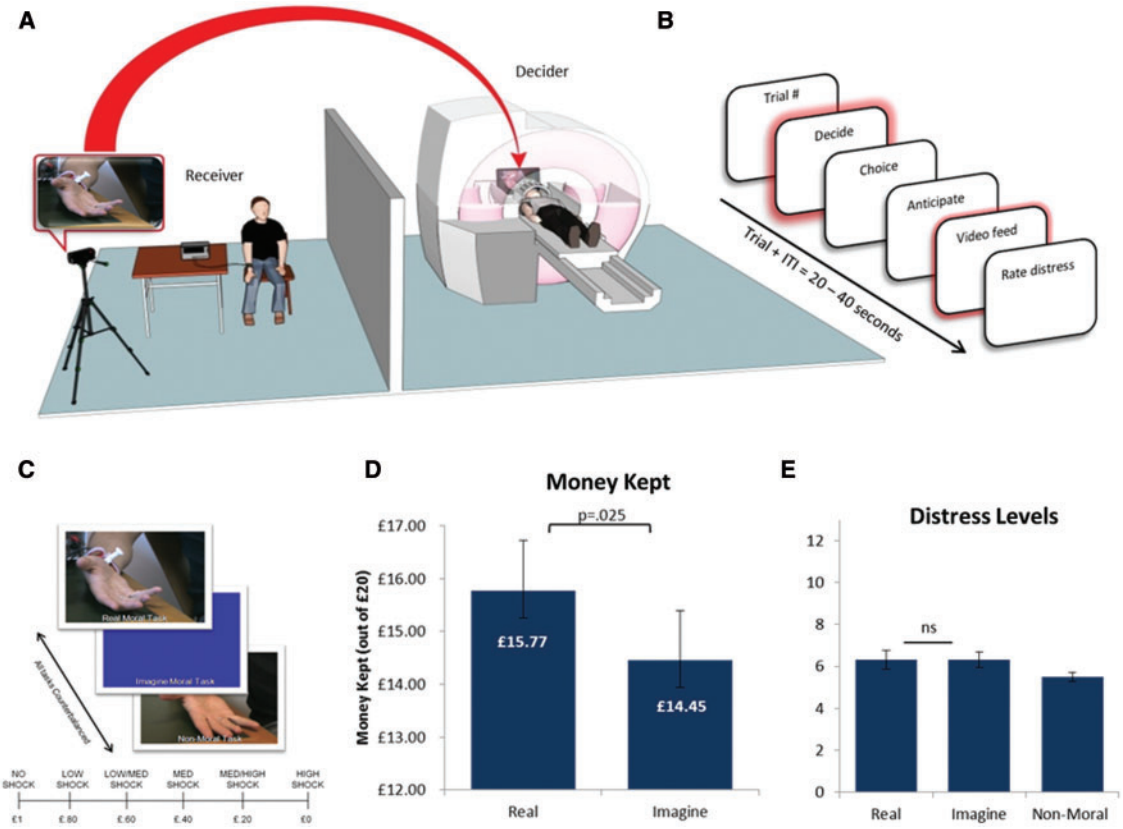

Fig. 1 Experimental setup, trial sequence (highlighting analyzed epochs) and behavioral data: (**A**) The Receiver (a confederate) sits in an adjoining testing laboratory to the scanning facility where the Decider (true subject) is undergoing fMRI. The Decider is told that any money left at the end of the task will be randomly multiplied up to 10 times, giving Deciders as much as £200 to take home. The Decider is also required to view, via prerecorded video feed, the administration of any painful stimulation to the Receiver, who is hooked up to an electric stimulation generator. (**B**) All three tasks (Real PvG, Imagine PvG and Non-Moral task) follow the same event-related design, with the same structure and timing parameters. Our analytical focus was on the Decide event (>11 s). The Video event (4 s), which was spaced a fixed 11 s after the Decide event, was also used in the analysis. (**C**) Still images of each task illustrating the video the Decider saw while in the scanner: Real PvG video, Imagine PvG video, and Non-Moral video, respectively. VAS scale Deciders used to indicate amount of money to give up/stimulation to deliver per trial. (**D**) Significantly more Money Kept in the Real PvG Task as compared to the Imagine PvG Task ($P = 0.025$; error bars = 1 S.E.M). (**E**) No significant differences between distress levels in response to the Video event across moral tasks.

chose to relinquish, the lower the painful stimulations inflicted on the Receiver, the key behavioral variable being how much money Deciders kept (with larger amounts indicating that personal gain was prioritized over Receiver's pain). The task comprised a series of eight screens per trial across 20 trials. Each trial began with a screen displaying the running amount of the subject's bank total (£20 on Trial 1) and current trial number. Subjects then had up to 11 s to decide upon and use a visual analogue scale (VAS) to select the amount of money they wanted to spend on that trial (up to £1) and thus the corresponding painful stimulation to be administered to the Receiver. This 11-s phase was partitioned into the 'Decide' and 'Select' periods. The Decide screen was presented for a fixed 3 s during which subjects were asked to think about their decision, so that when the select screen appeared, subjects could move the cursor to make their selection any time within the next 8 s. This design was used in order to introduce a variable jitter within the trial sequence. After making a selection, subjects saw a 3-s display of their choice before experiencing an 8-s anticipation phase—during which subjects were told their choice was being transmitted over the internal network to the other testing laboratory where the Receiver was connected to the electric stimulation generator. Following this anticipation period, subjects viewed a 4-s video of the stimulation being administered (Video event) to the Receiver, or no stimulation if they had opted to spend the full £1 permitted on a given trial. Subjects viewed a video feed of the Receiver's hand during stimulation administration. Finally, subjects used a 13-point VAS to rate their distress levels on viewing the consequences of their decision, before viewing a 4-s inter-trial-interval. At the conclusion of the 20 trials, subjects were able to press a button to randomly multiply any remaining money between 1 and 10 times, thus giving a maximum possible financial gain of £200. (See Supplementary Materials for descriptions of the Imagine PvG and Non-Moral tasks.)

Imaging methods

MRI scanning was conducted at the Medical Research Council Cognition and Brain Sciences Unit on a 3-Tesla Trio Tim MRI scanner by using a head coil gradient set. Whole-brain data were acquired with echoplanar T2*-weighted imaging (EPI), sensitive to BOLD signal contrast (48 sagittal slices, 3 mm thickness; Repetition Time (TR) = 2400 ms; Time to Echo (TE) = 30 ms; flip angle = 78°; Field of View (FOV) = 192 mm). To provide for equilibration effects, the first seven volumes were discarded. T1-weighted structural images were acquired at a resolution of $1 \times 1 \times 1$ mm. Statistical parametric mapping software was used to analyze all data. Pre-processing of fMRI data included spatial realignment, co-registration, normalization and smoothing. To control for motion, all functional volumes were re-aligned to the mean volume. Images were spatially normalized to standard space using the Montreal Neurological Institute (MNI) template with a voxel size of $3 \times 3 \times 3$ mm and smoothed using a Gaussian kernel with an isotropic full width at half maximum of 8 mm. In

addition, high-pass temporal filtering with a cutoff of 128 s was applied to remove low-frequency drifts in signal.

Statistical analysis

After pre-processing, statistical analysis was performed using the general linear model (GLM). Analysis was carried out to establish each participant's voxel-wise activation during the following events: making the decision of how much money to keep/which stimulations to administer (Decide event; Figure 1B) and watching the stimulation be administered (Video event; Figure 1B). Activated voxels were identified using an event-related statistical model representing each of the experimental events, convolved with a canonical hemodynamic response function and mean-corrected. Six head-motion parameters defined by the realignment were added to the model as regressors of no interest. For each fMRI experiment, contrast images for the Decide and Video events were calculated using GLMs and separately entered into full factorial analyses of variances (ANOVAs).

For group statistics, ANOVAs were used. For all three tasks (Real PvG, Imagine PvG and Non-Moral), the Decide event and the Video event were used in the following contrasts: (i) Real PvG > Imagine PvG, (ii) Imagine PvG > Real PvG and (iii) Real PvG > Non-Moral. A parametric regression analysis was used to explore which brain regions showed a correlation with Money Kept across the Real PvG task. We used a 1–6 parametric regressor weighted to the money chosen per trial—corresponding to the VAS scale used during the Decide event (Figure 1C). No significant activity was found for a parametric regression analysis for the Imagine PvG task. We report activity at $P < 0.001$ uncorrected for multiple spatial comparisons across the whole brain and $P < 0.05$ Family Wise Error (FWE) corrected for the following a priori regions of interest (ROIs; attained by independent coordinates): anterior insula, posterior cingulate cortex (PCC), medial and dorso-medial PFC (mPFC; dmPFC), hippocampus, temporoparietal junction (TPJ), amygdala and dorsolateral PFC (dlPFC). Coordinates were taken from previous related studies[1].

RESULTS

Behavioral results

Our study was motivated by the observation that moral action does not always reflect moral principle. Based on this, we anticipated that when the opportunity for making real money was salient, participants would favor financial self-interest (at the expense of the Receiver's pain) more during the real condition when compared with the hypothetical condition. This prediction was confirmed with subjects keeping significantly more money in the Real (£15.77, s.d. ±3.56) *vs* Imagine PvG task (£14.45, s.d. ±2.94; $t = 2.52$; $P = 0.025$; paired samples *t*-test, two-tailed; Figure 1D). Importantly, subjects showed no obvious strategy acquisition effects for keeping money over time (see Supplementary Analysis for details). There was no significant correlation between their ratings of the believability of the task and their behavioral performance (Money Kept), $r = -0.22$, $P > 0.1$. Furthermore, amount of Money Kept could not be explained by subjects modifying their decisions in response to reputation management or feelings of being watched (Landsberger, 1958; $r = 0.284$; $P = 0.325$, see Supplementary Methods for details). Self-reported distress ratings following the viewing of the Video event revealed that the Real PvG was no more distressing than imagining the painful stimulations in the Imagine PvG task ($t = 0.13$; $P = 0.89$; paired samples *t*-test, two-tailed; Figure 1E). This suggests that the emotional manipulation of watching an aversive video of the moral decision (when compared with viewing a blue screen and simulating the feedback of the decision) had no differential effect on participants' distress. There was, however, a significant difference between the distress levels reported in the Real PvG compared with the Non-Moral task ($t = -2.29$; $P = 0.039$; paired samples *t*-test, 2 tailed; Figure 1E).

[1]We used a priori coordinates to define ROI in our analysis. All ROIs were selected on the basis of independent coordinates using a sphere of 6-10 mm (sphere size was defined by the corresponding structure) and corrected at $P < 0.05$ FWE and were attained through MarsBaRs. Peak voxels are presented in the tables at $P < 0.001$ uncorrected and images are shown at $P < 0.005$ uncorrected.

Table 1 Decide event of Real PvG contrasted to Non-Moral task (Real PvG Decide > Non-Moral PvG Decide)

Region	Peak MNI coordinates			*z* value
Right ACC	14	38	28	3.12
Left amygdala	−26	−2	−26	3.00
Right amygdala	28	−8	−28	3.00
Right fusiform	28	−64	−10	3.49
A priori ROIs	**MNI coordinates**			***t*-statistic**
Right amygdala[a]	28	−4	−26	3.61
Left amygdala[a]	−20	−6	−26	3.39

ROI = regions of interest with 6 mm sphere corrected at $P < 0.05$ FWE using a priori independent coordinates from previous study: [a]Akitsuki and Decety (2009).

Imaging results

Real moral vs non-moral decisions

In line with the traditional research (Greene et al., 2001), we first compared moral decisions in the Real PvG to decisions in the Non-Moral task, which revealed bilateral amygdala and anterior cingulate cortex (ACC; the Decide event in the Real PvG contrasted with the Decide event in the Non-Moral Task [Table 1])—two regions that are known to process emotionally aversive stimuli (Bechara et al., 2003), especially during emotional conflict (Etkin et al., 2011). That decisions made during the Real PvG reveal patterns of activation within emotion processing areas likely reflects the fact that moral decisions are more emotionally arousing than decisions made within a non-moral context.

Real and hypothetical decisions

To specifically elucidate the differences between real and hypothetical moral decisions, we compared the Decide event (Figure 1B) for the Imagine and Real PvG tasks, highlighting the brain regions distinct to each condition. Significant activation in the PCC, bilateral hippocampus and posterior parietal lobe—all regions essential in imagination and prospection (Schacter et al., 2007)—were greater for hypothetical moral decisions (Figure 2A). Applying a priori ROIs derived from research on the brain's construction system (Hassabis and Maguire, 2009) revealed a remarkably shared neural system with hypothetical moral decisions (Table 2). Additional a priori ROIs drawn from the moral literature—mPFC and dlPFC (Greene et al., 2001)—also showed greater activation for imagined moral choices. Parameter estimates of the beta values for these ROIs confirmed that these regions were more sensitive to hypothetical moral decisions, relative to real moral decisions (Figure 2A). In contrast, activation in the bilateral ventral TPJ [BA 37], bilateral amygdala, putamen and ACC were more active for real moral decisions (Figure 2B; Table 3). As with the previous contrast, we first applied a priori ROIs and then examined the parameter estimates to ensure that the amygdala and TPJ were significantly more active during real moral decisions. These regions are well documented within the social neuroscience literature and have been closely associated with processing stimuli with emotional and social significance (Phelps, 2006).

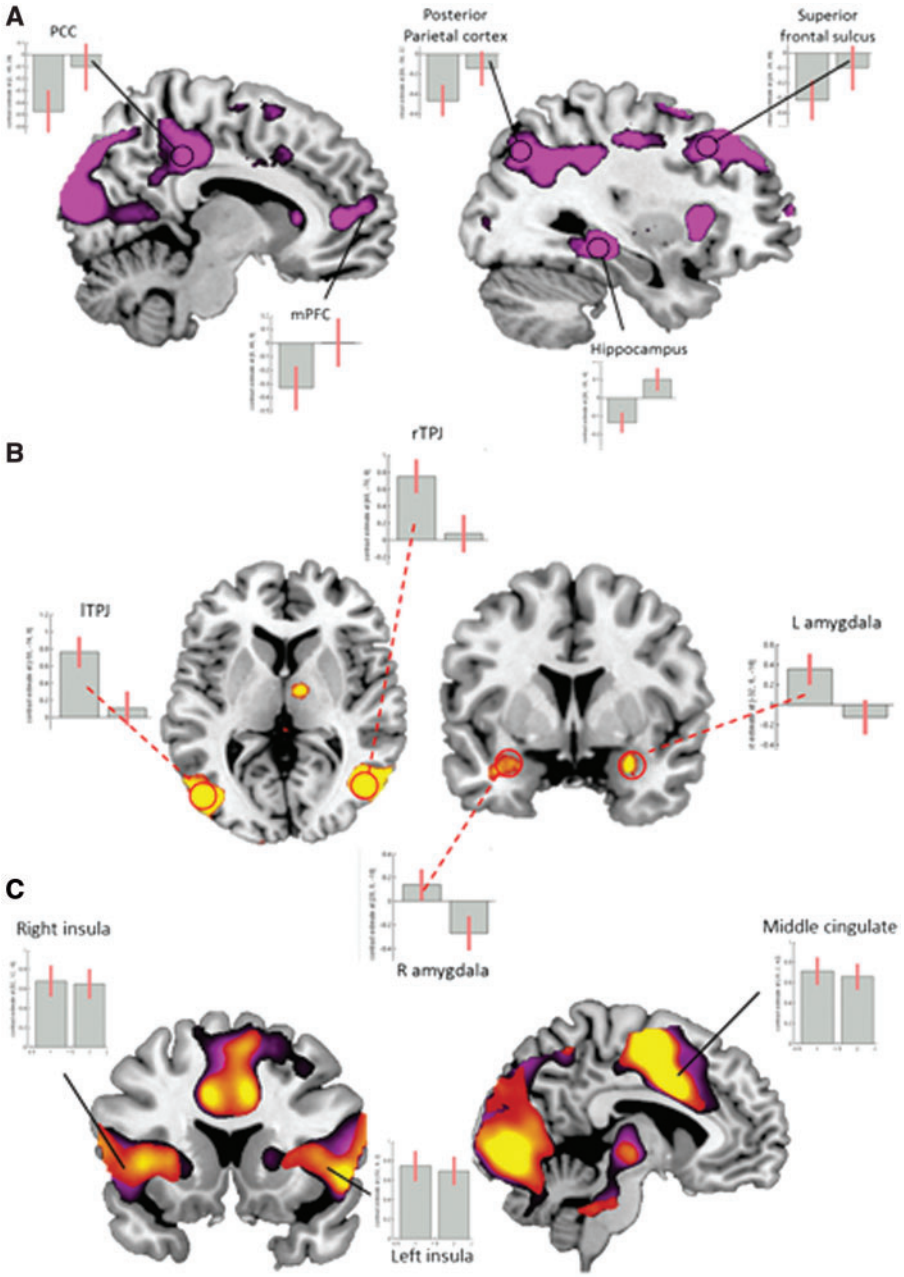

Fig. 2 Real and Imagine Moral networks: (**A**) Imagine Moral Network: Comparing the Imagine PvG Decide event > Real PvG Decide event reveals significant activation in the PCC, mPFC, posterior parietal cortex, superior frontal sulcus and hippocampus. A priori ROIs (indicated by circles and corrected at $P < 0.05$ FWE) and parameter estimates reveal that hypothetical moral decisions map closely onto the brain's construction system. (**B**). Real Moral Network: Contrasting the Decide event of the Real PvG > Imagine PvG activates bilateral TPJ and amygdala. A priori ROIs and parameter estimates for these regions were found to be more significant during the Real decision than during the Imagine decision. (**C**). Shared Moral Network: A conjunction analysis of Real and Imagine moral decisions reveals robust activation in the empathy for pain matrix, and parameter estimates of the middle cingulate and bilateral insula illustrate comparable activations for both conditions. All coordinates in MNI space and results portrayed on sections of the mean structural scan at $P < 0.005$ uncorrected. Both whole brain analysis ($P < 0.001$ uncorrected) and a priori regions of interest (FWE $P < 0.05$) were used for all contrasts. A complete list of activated areas and ROIs can be found in Tables 2–4.

Shared moral networks

We ran a conjunction analysis of all moral decisions to determine if there is a common neural circuitry between real and hypothetical moral decisions (real moral decisions compared with non-moral decisions, along with imagined moral decisions compared with non-moral decisions (Real PvG Decide + Imagine PvG Decide). The results revealed that moral decisions, regardless of condition, shared common activation patterns in the bilateral insula (extending posterior to anterior), middle cingulate (MCC), bilateral dlPFC and bilateral TPJ extending into the posterior superior temporal sulcus (BA 40—which differs from the peak coordinates found for real moral decisions; Figure 2C; Table 4).

Real vs imagine feedback

Although subjects' distress ratings across moral tasks were not significantly different [$F(1) < 1$, $P = 0.99$ (Figure 1E)], we wanted to first ensure that the video feedback event in the Real PvG was not driving activation during the Decision event and then examine the Deciders'

Table 2 Decide event of Imagine PvG contrasted to Real PvG (Imagine PvG Decide > Real Decide)

Region	Peak MNI coordinates			*z* value
Right hippocampus	34	−30	−4	5.70
Left hippocampus	−32	−18	−10	3.80
Right posterior parietal cortex	42	−66	38	5.39
Right occipital lobe	6	−94	24	5.45
Right PCC	8	−32	38	4.10
rACC/MFG	4	44	6	5.02
Right mid temporal lobe	64	−38	−10	5.10
Left mid temporal lobe	−60	−48	−6	4.48
Left dlPFC	−18	32	42	4.26
MCC	−8 46	−16	46	4.08
Left caudate	−18	−10	20	3.95
Right putamen	28	18	6	5.20
A priori ROIs	**MNI coordinates**			***t*-statistic**
vmPFC[a]	3	24	−9	3.78
Right superior frontal sulcus[a]	27	27	45	3.78
Right hippocampus[a]	21	−24	−12	6.81
Left parahippocampus gyrus[a]	−18	−35	−15	7.80
Right parahippocampus gyrus[a]	0 33	−42	−12	6.77
Left posterior parietal cortex[a]	−48	−78	24	3.80
Right posterior parietal cortex[a]	45	−55	24	4.56
mPFC[b]	1	53	22	3.80
dlPFC[b]	44	36	28	5.08
Left angular gyrus[b]	−48	−68	25	4.09
Right angular gyrus[b]	50	−60	28	4.33
PCC[b]	−4	−57	36	4.13

ROI = regions of interest corrected at $P < 0.05$ FWE using a priori independent coordinates from previous studies: [a]Hassabis et al. (2007); [b]Greene et al. (2001).

Table 3 Decide event of Real PvG contrasted to Imagine PvG (Real PvG Decide > Imagine Decide)

Region	Peak MNI coordinates			*z* value
Left TPJ	−44	−74	0	6.58
Right TPJ	46	−68	2	6.44
dlPFC	54	22	6	4.62
SMA	46	−18	62	4.19
Left amygdala	−30	10	−18	4.15
Right amygdala	26	10	−18	4.11
ACC	16	40	26	3.98
Thalmus/STA region	10	−12	10	3.97
Right anterior insula	28	32	−8	3.23
A priori ROIs	**MNI coordinates**			***t*-statistic**
Left TPJ[a]	−53	−71	6	10.34
Right TPJ[a]	50	−75	9	10.43
Right amygdala[b]	28	−4	−26	4.85
Left amygdala[b]	−20	−6	−26	3.32

ROI = regions of interest corrected at $P < 0.05$ FWE using a priori independent coordinates from previous studies: [a]Borg et al. (2006); [b]Akitsuki and Decety (2009).

Table 4 Conjunction analysis of all moral decisions (Real PvG Decide + Imagine PvG Decide)

Region	Peak MNI coordinates			*z* value
Visual cortex	10	−86	16	7.55
Right insula	52	12	−6	5.84
Left insula	−52	12	2	4.51
Right TPJ	64	−36	30	5.15
Left TPJ	−60	−33	20	4.02
Mid cingulate	−8	2	42	6.67
Right dlPFC	32	44	28	3.91
Left dlPFC	−34	42	26	3.92
A priori ROIs	**MNI coordinates**			***t*-statistic**
Right anterior insula[a]	60	15	3	7.27
Left anterior insula[a]	−48	12	2	7.85
ACC[a]	−9	6	42	8.47
Left TPJ[b]	−53	−71	6	10.34
Right TPJ[b]	50	−75	9	4.60
Right TPJ/Angular gyrus[c]	50	−56	20	3.59

ROI = regions of interest corrected at $P < 0.05$ FWE using a priori independent coordinates from previous studies: [a]Singer et al. (2004); [b]Borg et al (2006); [c]Greene et al. (2001).

Table 5 Video event of Real PvG contrasted to Imagine PvG (Real PvG Video > Imagine PvG video)

Region	Peak MNI coordinates			*z* value
Right TPJ	52	−58	4	5.15
Left TPJ	−52	−68	4	5.34
Right anterior insula	50	18	10	4.60
Right anterior insula	38	24	−6	4.05
Left anterior insula	−28	12	−18	3.54
Left anterior insula	−33	20	−6	3.00
Mid cingulate	8	32	44	3.38
A priori ROIs	**MNI coordinates**			**t-statistic**
Right anterior insula[a]	60	15	3	5.72
Right anterior insula[a]	39	12	3	4.53
Right anterior insula[a]	42	27	−6	4.84

ROI = regions of interest corrected at $P < 0.05$ FWE using a priori independent coordinates from previous study; [a]Singer et al. (2004).

socio-affective engagement with the task. Accordingly, we compared neural activation during the Video feed event (Figure 1B) and predicted greater anterior insula activation indexing empathy for pain (Singer et al., 2004) in the Real PvG when compared with the Imagine PvG. As we expected, a comparison analysis of the Video event for the Real *vs* Imagine PvG (Table 5) revealed greater bilateral anterior insula activation when participants watched real administrations of painful stimulations. This further supports the well-documented proposal that the anterior insula codes for the empathic experience of another's pain (Singer et al., 2004) and suggests that participants experienced greater socio-affective engagement in the Real version of the task.

Self-interested vs pro-social behavior

One strength of the use of multiple trials in the PvG task is that it gives Deciders the option to either maintain a purely black (keep £20; maximize shocks) or white (keep £0; remove shocks) moral stance, or to position themselves somewhere within the moral 'gray area' (£0 < keep < £20). This not only has ecological validity, reflecting people's tendency to qualify moral decisions but also allows us to investigate brain regions associated with different shades of 'moral gray.' We therefore conducted a parametric regression analysis and found that increasingly self-interested behavior on the Real PvG task (when Deciders kept more money, parametrically weighted on a scale from 1 to 6) was associated with increased activity in the dorsal ACC (dACC), bilateral dlPFC and orbital frontal cortex (OFC; Figure 3A;

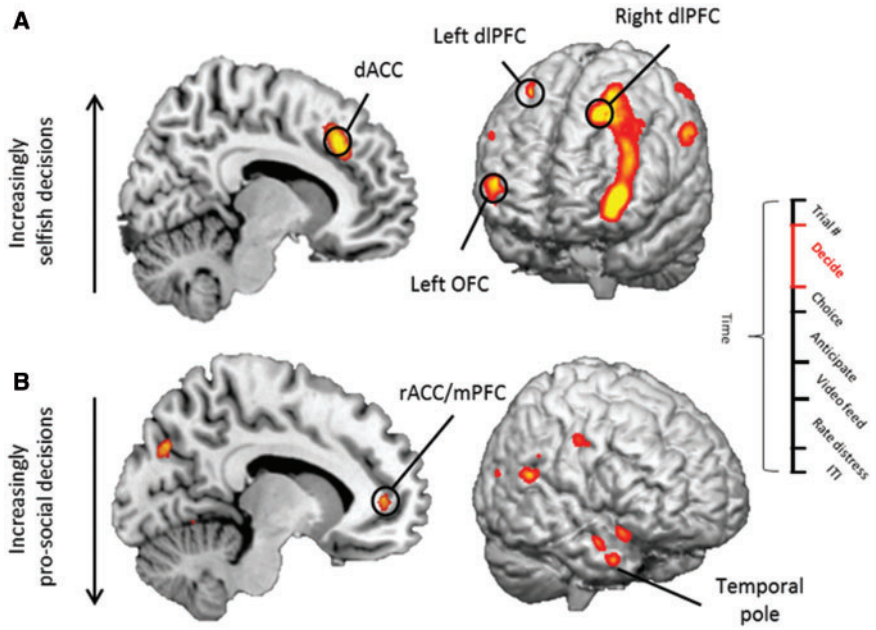

Fig. 3 Dissociable networks for real selfish and pro-social moral decisions (**A**) Parametric regression analysis (trial-by-trial) of the Decide event of the Real PvG for increasingly selfish behaviors (greater Money Kept) activates the dACC, bilateral OFC and bilateral dlPFC. A priori ROIs (indicated by circles and corrected at $P < 0.05$ FWE) were found to be significantly activated for these regions. (**B**) Parametric regression analysis of the Real PvG Decide event for increasingly pro-social decisions (greater money given up) reveals significant activation in the rostral ACC/mPFC, right temporal pole and right anterior insula. An a priori ROI for the rACC corrected at $P < 0.05$ FWE was found to be significantly activated. All results portrayed on both axial sections and rendered images at $P < 0.005$ uncorrected. Both whole brain analysis ($P < 0.001$ uncorrected) and a priori regions of interest (FWE $P < 0.05$) were used for all contrasts. All coordinates in MNI space: a complete list of activated areas and ROIs can be found in Tables 6 and 7.

Table 6 Parametric modulation weighted by monetary choice for Real PvG Selfish Decisions (Real PvG Decide, weighted 1–6)

Region	Peak MNI coordinates			*z* value
Right mid frontal gyrus/dlPFC	28	38	48	3.62
Right inferior OFC	40	46	0	3.58
dACC	10	26	38	3.25
Left mid frontal gyrus	−32	12	28	3.16
Left dlPFC	−30	8	52	3.47
Right dlPFC	28	10	52	3.20
A priori ROIs	**MNI coordinates**			***t*-statistic**
dmPFC[a]	0	24	40	3.61
Middle frontal gyrus[a]	−24	2	52	4.04
Left frontal pole[a]	−36	50	10	3.39
ACC[a]	6	24	34	3.72

ROI = regions of interest corrected at $P < 0.05$ FWE using a priori independent coordinates from previous study; [a]Liu et al. (2011)

Table 7 Parametric modulation weighted by monetary choice for Real PvG Pro-Social Decisions (Real PvG Decide, weighted 1–6)

Region	Peak MNI coordinates			*z* value
Right cuneus	14	-82	28	3.96
MFG/rACC	−12	46	6	3.00
Left TPJ	−64	−38	20	3.00
Left temporal pole	−48	14	30	3.00
Left anterior insula	−38	12	−12	2.85
A priori ROIs	**MNI coordinates**			***t*-statistic**
MFG/rACC[a]	−16	49	9	3.33

ROI = regions of interest corrected at $P < 0.05$ FWE using a priori independent coordinates from previous study: [a]Takahashi et al. (2004).

Table 6), regions sensitive to cognitive control (Ochsner and Gross, 2005) reward (Kringelbach, 2005) and, in particular, monetary gain (O'Doherty et al., 2001). Activations associated with decreasing self-interest (pro-social behavior) were predominately localized to the mPFC/rostral ACC (rACC), temporal pole and anterior insula (Figure 3B; Table 7; parametric regressions run on the Imagine PvG did not reveal any similar regions). Critically, these parametric analyses also clarified that the activations found during real moral decisions (regardless of the nature of the decision) were not simply an effect of reward classification for monetary gain or loss.

Individual differences

To test for potential motivating factors driving the selfish and pro-social behavior found in the Real PvG, we explored individual differences. Using post-scan questionnaires scores as covariates in a correlated regression for the Real PvG Decision event revealed differential activations for empathic concern and perspective taking (Davis, 1983) and self-reported similarity ratings. Subgenual ACC correlated with increased empathic concern (Table 8) while decreasing empathic concern and perspective-taking activated left putamen, dACC, bilateral dlPFC and bilateral OFC (Table 9). Finally, similarity ratings negatively correlated with the amount of Money Kept and elicited activation in the right anterior insula, while increasing similarity ratings correlated with activation in the ACC mPFC, dmPFC and left OFC (Tables 10 and 11, respectively).

DISCUSSION

This study examined the moral dynamic of self-gain *vs* other-welfare during real and hypothetical conditions. Our behavioral results show that moral decisions with real consequences diverge from hypothetical

Table 8 Correlation regression for increasing empathic concern (Real PvG Decide > Imagine PvG Decide)

Region	Peak MNI coordinates			*z* value
Subgenal ACC	2	28	−2	3.15
A priori ROIs	MNI coordinates			*t*-statistic
Subgenal ACC[a]	6	36	−4	3.82

ROI = regions of interest corrected at $P < 0.05$ FWE using a priori independent coordinates from previous study: [a]Zahn et al. (2009).

Table 9 Correlation regression for decreasing empathic concern and perspective taking (Real PvG Decide > Imagine PvG Decide)

Region	Peak MNI coordinates			*z* value
Left superior temporal sulcus	−38	−74	44	4.12
Right superior temporal sulcus	30	−74	48	4.12
Left putamen	−14	10	2	3.45
dACC	−8	36	34	3.30
Right dlPFC	32	6	46	3.27
Left dlPFC	−32	4	54	3.22
Left OFC	−24	42	2	4.70
Right OFC	30	58	8	4.03
Right dlPFC	30	24	48	3.65
Left dlPFC	−24	16	52	3.62
mPFC	16	50	4	3.72

Table 10 Correlation regression for decreasing similarity ratings (Real PvG Decide > Imagine PvG Decide)

Region	Peak MNI coordinates			*z* value
Right anterior insula	44	28	0	3.19

Table 11 Correlation regression for increasing similarity ratings (Real PvG Decide > Imagine PvG Decide)

Region	Peak MNI coordinates			*z* value
Left middle frontal gyrus	−28	18	44	3.40
MPFC/rACC	4	38	−2	3.01
Left Hippocampus	−30	−40	−4	3.16

moral choices, verifying the 'hypothetical bias' effect (Kang et al., 2011). Compared with imagining their moral actions, people who make moral decisions under real conditions keep more money and inflict more pain on another subject. Although the research exploring real moral action is limited (Moll et al., 2006; Baumgartner et al., 2009; Greene and Paxton, 2009), our results stand in stark contrast to findings demonstrating that people act more morally than they think they will (Teper et al., 2011). Our results also contradict the accumulated research illustrating a basic aversion to harming others (Greene et al., 2001; Cushman et al., 2012). We contend that this is likely due to the fact that many of the moral scenarios used within the moral literature do not pit the fundamental motivation of not harming others (physically or psychological) against that of maximizing self-gain (Haidt, 2007). Accordingly, our findings reveal that engaging the complex motivations of self-benefit—a force endemic to many moral decisions—can critically influence moral action.

Our fMRI results identify a common neural network for real and hypothetical moral cognition, as well as distinct circuitry specific to real and imagined moral choices. Moral decisions—regardless of condition—activated the insula, MCC and dorsal TPJ, areas essential in higher order social processes, such as empathy (Singer et al., 2004). This neural circuitry is well instantiated in the social neuroscience literature and fits with the findings that moral choices are influenced by neural systems whose primary role is to facilitate cooperation (Rilling and Sanfey, 2011). The TPJ has been specifically implicated in decoding social cues, such as agency, intentionality and the mental states of others (Young and Saxe, 2008). For example, TPJ activation correlates with the extent to which another's intentions are taken into account (Young and Saxe, 2009) and transiently disrupting TPJ activity leads to interference with using mental state information to make moral judgments (Young et al., 2010). Although there is a large amount of research indicating that the TPJ codes for our ability to mentalize, there is also evidence that the TPJ activates during attentional switching (Mitchell, 2008). In addition, one study revealed that patients with lesions to the TPJ do not show domain-specific deficits for false belief tasks (Apperly et al., 2007). Although these differential findings suggest that the specific functionality of the TPJ remains unclear, we propose that TPJ engagement during real and imagined moral decisions suggests a similar mentalizing process is at play in both real and hypothetical moral decision-making: when deciding how much harm to apply to another, subjects may conscript a mental state representation of the Receiver, allowing them to weigh up the potential consequences of their decision. This neural finding reinforces the role of the TPJ—and thus the likely role of mental state reasoning and inference—in moral reasoning.

However, we also found distinct neural signatures for both real and imagined moral decisions. In line with the literature, hypothetical moral decisions were specifically subserved by activations in the PCC and mPFC—regions also implicated in prospection, by which abridged simulations of reality are generated (Gilbert and Wilson, 2007). Although the overall pattern of brain activation during these hypothetical moral decisions replicates the moral network identified in previous research (Greene et al., 2001), the fact that the PCC and mPFC are activated both during prospection and during hypothetical moral decision-making implies that this region is recruited for a wide spectrum of imagination-based cognition (Hassabis and Maguire, 2009). Thus, either hypothetical moral decisions and imagination share a similar network or hypothetical moral decisions significantly rely on the imperfect systems of prospection and imagination. Further research exploring whether the PCC and mPFC are specific to hypothetical moral decisions, or recruited more generally for imagining future events, would help clarify their roles within the moral network.

In contrast, real moral decisions differentially recruited the amygdala. These results are consistent with the vast literature implicating the amygdala in processing social evaluations (Phelps, 2006), emotionally relevant information (Sander et al., 2003) and salient stimuli (Ewbank et al., 2009). Research on moral cognition further implicates amygdala activation in response to aversive moral phenomena (Berthoz et al., 2006; Kedia et al., 2008; Glenn et al., 2009); however, this finding is not systematically observed in moral paradigms (Raine and Yang, 2006). In line with the literature, it is possible that in the Real PvG task the amygdala is coding the aversive nature of the moral decision; however, distress ratings indicated that both conditions were perceived as equally aversive. Accordingly, an alternative interpretation is that the amygdala is monitoring the salience, relevance and motivational significance (Mitchell et al., 2002) of the real moral choice space.

Decisions, which produce real aversive consequences (i.e. lose money or harm another), are far more salient and meaningful than decisions that do not incur behaviorally relevant outcomes. The amygdala is also commonly recruited for decisions which rely on social signals to emotionally learn positive and negative associations (Hooker et al., 2006). It is possible that the amygdala activation found for real moral decisions is signaling reinforcement expectancy information of both the positively (self-benefit) and negatively (harm to another) valenced stimuli (Blair, 2007), which then subsequently guides behavior (Prevost et al., 2011). This theory not only accounts for the differential behavioral findings between the real and hypothetical conditions but also it is consistent with the more general theoretical consensus regarding human moral cognition (Moll et al., 2005), which emphasizes how lower order regions like the amygdala modulate higher order rational processes (Dalgleish, 2004).

Our fMRI results further indicate that there are dissociable neural mechanisms underlying selfish and pro-social decisions. In the Real PvG, decisions that maximized financial benefit (selfish decisions) correlated with activity in the OFC, dlPFC and dACC—regions that support the integration of reward and value representations (Schoenbaum and Roesch, 2005), specifically monetary gain (Holroyd et al., 2004) and loss (Bush et al., 2002). Furthermore, the dACC was found to negatively correlate with empathic concern scores and positively correlate with self-reported similarity ratings in the Real PvG task. Together, this suggests that the dACC may be monitoring conflicting motive states (Etkin et al., 2011). However, the dACC has been further implicated in a variety of other functions, including emotion regulation (Etkin et al., 2011), and weighing up different competing choices (Mansouri et al., 2009). Thus, it is equally plausible that the dACC is processing the conflicting negative emotions involved with choosing to harm another for self-gain (Amodio and Frith, 2006).

In the PvG task, the morally guided choice is to give up the money to prevent harm to another. Unlike selfish decisions, such pro-social decisions showed significantly greater activation in the rACC/mPFC and right temporal pole, demonstrating that the nature of real moral decisions can be predicted by dissociable networks within the PFC. The rACC/mPFC is a structure engaged in generating empathic feelings for in-group members (Mathur et al., 2010) and for coding feelings of altruistic guilt and distress during theory of mind tasks (Fletcher et al., 1995). Clinical data have also shown that lesions to this area stunt moral emotions, such as compassion, shame and guilt, and contribute to overall deficits in emotional processing (Mendez and Shapira, 2009). In fact, research has demonstrated the rACC/mPFC as a region that responds specifically to the aversion of not harming others (Young and Dungan, 2011). Based on this, we propose that the rACC/mPFC activation found for pro-social decisions could be attributed to the empathic response generated by the emotional aversion (distress) of harming another—a key motivational influence and proximate mechanism of altruistic behavior.

Theorists have pointed to the importance of studying moral cognition in ecological valid and consequence-driven environments (Casebeer, 2003; Moll et al., 2005). Our results illustrate that specific regions of the moral network subserve moral choices—regardless of whether they are real or imagined. However, we also found a divergence between real moral behavior and hypothetical moral intentions—which was reflected in the recruitment of differential neurobiological systems. Thus, if morality is a domain where situational influences and the impact of imminent, real consequences can sway our decisions, then it is crucial that cognitive neuroscience investigate moral decision-making under real conditions. This seems especially relevant in light of this new neurobiological evidence, supporting what the philosopher Hume presciently noted—'the most lively thought is still inferior to the dullest sensation' (Hume, 1977).

SUPPLEMENTARY DATA

Supplementary data are available at *SCAN* online.

FUNDING

This research was supported by the Medical Research Council Cognition and Brain Sciences Unit.

REFERENCES

Ajzen, I., Brown, T.C., Carvajal, F. (2004). Explaining the discrepancy between intentions and actions: the case of hypothetical bias in contingent valuation. *Personality and Social Psychology Bulletin*, *30*(9), 1108–21.

Akitsuki, Y., Decety, J. (2009). Social context and perceived agency affects empathy for pain: an event-related fMRI investigation. *NeuroImage*, *47*(2), 722–34.

Amodio, D.M., Frith, C.D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nature reviews. Neuroscience*, *7*(4), 268–77.

Apperly, I.A., Samson, D., Chiavarino, C., Bickerton, W.L., Humphreys, G.W. (2007). Testing the domain-specificity of a theory of mind deficit in brain-injured patients: evidence for consistent performance on non-verbal, "reality-unknown" false belief and false photograph tasks. *Cognition*, *103*(2), 300–21.

Baumgartner, T., Fischbacher, U., Feierabend, A., Lutz, K., Fehr, E. (2009). The neural circuitry of a broken promise. *Neuron*, *64*(5), 756–70.

Bechara, A., Damasio, H., Damasio, A.R. (2003). Role of the amygdala in decision-making. *Annals of the New York Academy of Sciences*, *985*, 356–69.

Berthoz, S., Grezes, J., Armony, J.L., Passingham, R.E., Dolan, R.J. (2006). Affective response to one's own moral violations. *Neuroimage*, *31*(2), 945–50.

Blair, R.J. (2007). The amygdala and ventromedial prefrontal cortex in morality and psychopathy. *Trends in cognitive sciences*, *11*(9), 387–92.

Borg, J., Hynes, C., Van Horn, J., Grafton, S., Sinnott-Armstrong, W. (2006). Consequences, action, and intention as factors in moral judgments: an fMRI study. *Journal of cognitive Neuroscience*, *5*, 803–17.

Bush, G., Vogt, B.A., Holmes, J., et al. (2002). Dorsal anterior cingulate cortex: a role in reward-based decision making. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(1), 523–8.

Casebeer, W.D. (2003). Moral cognition and its neural constituents. *Nature reviews. Neuroscience*, *4*(10), 840–6.

Cushman, F., Gray, K., Gaffey, A., Mendes, W.B. (2012). Simulating murder: the aversion to harmful action. *Emotion*, *12*(1), 2–7.

Dalgleish, T. (2004). The emotional brain. *Nature Reviews Neuroscience*, *5*(7), 583–9.

Davis, M.H. (1983). Measuring individual differences in empathy: evidence for a multidimensional approach. *Journal of Personality and Social Psuchology*, *44*(1), 113–26.

Etkin, A., Egner, T., Kalisch, R. (2011). Emotional processing in anterior cingulate and medial prefrontal cortex. *Trends in cognitive sciences*, *15*(2), 85–93.

Ewbank, M.P., Barnard, P.J., Croucher, C.J., Ramponi, C., Calder, A.J. (2009). The amygdala response to images with impact. *Social cognitive and affective neuroscience*, *4*(2), 127–33.

FeldmanHall, O., Mobbs, D., Evans, D., Hiscox, L., Navardy, L., Dalgleish, T. (2012). What we say and what we do: the relationship between real and hypothetical moral choices. *Cognition*, *123*, 434–41.

Fletcher, P.C., Happe, F., Frith, U., et al. (1995). Other minds in the brain: a functional imaging study of "theory of mind" in story comprehension. *Cognition*, *57*(2), 109–128.

Gilbert, D.T., Wilson, T.D. (2007). Prospection: experiencing the future. *Science*, *317*(5843), 1351–4.

Glenn, A.L., Raine, A., Schug, R.A. (2009). The neural correlates of moral decision-making in psychopathy. *Molecular Psychiatry*, *14*(1), 5–6.

Greene, J.D., Paxton, J.M. (2009). Patterns of neural activity associated with honest and dishonest moral decisions. *Proceedings of the Academy of Sciences U S A*, *106*(30), 12506–11.

Greene, J.D., Sommerville, R.B., Nystrom, L.E., Darley, J., Cohen, J.D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*(5537), 2105–8.

Haidt, J. (2007). The new synthesis in moral psychology. *Science*, *316*(5827), 998–1002.

Hassabis, D., Kumaran, D., Maguire, E.A. (2007). Using imagination to understand the neural basis of episodic memory. *Journal of Neuroscience*, *27*(52), 14365–74.

Hassabis, D., Maguire, E.A. (2009). The construction system of the brain. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1521), 1263–71.

Higgins, E.T. (1987). Self-discrepancy: a theory relating self and affect. *Psychological review*, *94*(3), 319–340.

Holroyd, C.B., Nieuwenhuis, S., Yeung, N., et al. (2004). Dorsal anterior cingulate cortex shows fMRI response to internal and external error signals. *Nature Neuroscience*, *7*(5), 497–8.

Hooker, C.I., Germine, L.T., Knight, R.T., D'Esposito, M. (2006). Amygdala response to facial expressions reflects emotional learning. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *26*(35), 8915–22.

Hume, D., editor (1977) *An Equiry Concerning Human Understanding*, 2nd edn. Indianapolis: Hackett Publishing Company.

Kang, M.J., Rangel, A., Camus, M., Camerer, C.F. (2011). Hypothetical and real choice differentially activate common valuation areas. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *31*(2), 461–8.

Kedia, G., Berthoz, S., Wessa, M., Hilton, D., Martinot, J.L. (2008). An agent harms a victim: a functional magnetic resonance imaging study on specific moral emotions. *Journal of Cognitive Neuroscience*, *20*(10), 1788–98.

Krebs, D., Denton, K., Wark, G. (1997). The forms and functions of real-life moral decision-making. *Journal of Moral Education*, *26*(2), 131–45.

Kringelbach, M.L. (2005). The human orbitofrontal cortex: linking reward to hedonic experience. *Nature Reviews Neuroscience*, *6*(9), 691–702.

Landsberger, H. (1958). *Hawthorne Revisited: Management and the Worker, Its Critics, and Developments in Human Relations in Industry.* Ithaca, NY: Distribution Center, N.Y.S. School of Industrial and Labor Relations, Cornell University.

Liu, X., Hairston, J., Schrier, M., Fan, J. (2011). Common and distinct networks underlying reward valence and processing stages: a meta-analysis of functional neuroimaging studies. *Neuroscience and Biobehavioral Reviews*, *35*(5), 1219–36.

Mansouri, F.A., Tanaka, K., Buckley, M.J. (2009). Conflict-induced behavioural adjustment: a clue to the executive functions of the prefrontal cortex. *Nature Reviews. Neuroscience*, *10*(2), 141–52.

Mathur, V.A., Harada, T., Lipke, T., Chiao, J.Y. (2010). Neural basis of extraordinary empathy and altruistic motivation. *NeuroImage*, *51*(4), 1468–75.

Mendez, M.F., Shapira, J.S. (2009). Altered emotional morality in frontotemporal dementia. *Cognitive Neuropsychiatry*, *14*(3), 165–79.

Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal Psychology*, *67*, 371–8.

Mitchell, D.G., Colledge, E., Leonard, A., Blair, R.J. (2002). Risky decisions and response reversal: is there evidence of orbitofrontal cortex dysfunction in psychopathic individuals? *Neuropsychologia*, *40*(12), 2013–22.

Mitchell, J.P. (2008). Activity in right temporo-parietal junction is not selective for theory-of-mind. *Cerebral Cortex*, *18*(2), 262–71.

Moll, J., Krueger, F., Zahn, R., Pardini, M., de Oliveira-Souza, R., Grafman, J. (2006). Human fronto-mesolimbic networks guide decisions about charitable donation. *Proceedings of the Academy of Sciences U S A*, *103*(42), 15623–8.

Moll, J., Zahn, R., de Oliveira-Souza, R., Krueger, F., Grafman, J. (2005). Opinion: the neural basis of human moral cognition. *Nature Reviews Neuroscience*, *6*(10), 799–809.

O'Doherty, J., Kringelbach, M.L., Rolls, E.T., Hornak, J., Andrews, C. (2001). Abstract reward and punishment representations in the human orbitofrontal cortex. *Nature Neuroscience*, *4*(1), 95–102.

Ochsner, K.N., Gross, J.J. (2005). The cognitive control of emotion. *Trends in Cognitive Sciences*, *9*(5), 242–9.

Phelps, E.A. (2006). Emotion and cognition: insights from studies of the human amygdala. *Annual Review of Psychology*, *57*, 27–53.

Prevost, C., McCabe, J.A., Jessup, R.K., Bossaerts, P., O'Doherty, J.P. (2011). Differentiable contributions of human amygdalar subregions in the computations underlying reward and avoidance learning. *The European Journal of Neuroscience*, *34*(1), 134–45.

Raine, A., Yang, Y. (2006). Neural foundations to moral reasoning and antisocial behavior. *Social Cognitive and Affective Neuroscience*, *1*(3), 203–13.

Rilling, J.K., Sanfey, A.G. (2011). The neuroscience of social decision-making. *Annual Review of Psychology*, *62*, 23–48.

Sander, D., Grafman, J., Zalla, T. (2003). The human amygdala: an evolved system for relevance detection. *Reviews in the Neurosciences*, *14*(4), 303–16.

Schacter, D.L., Addis, D.R., Buckner, R.L. (2007). Remembering the past to imagine the future: the prospective brain. *Nature Reviews Neuroscience*, *8*(9), 657–661.

Schoenbaum, G., Roesch, M. (2005). Orbitofrontal cortex, associative learning, and expectancies. *Neuron*, *47*(5), 633–6.

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R.J., Frith, C.D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science*, *303*(5661), 1157–62.

Takahashi, H., Yahata, N., Koeda, M., Matsuda, T., Asai, K., Okubo, Y. (2004). Brain activation associated with evaluative processes of guilt and embarrassment: an fMRI study. *NeuroImage*, *23*(3), 967–74.

Teper, R., Inzlicht, M., Page-Gould, E. (2011). Are we more moral than we think?: exploring the role of affect in moral behavior and moral forecasting. *Psychological Science*, *22*(4), 553–8.

Wilson, T.D., Wheatley, T., Meyers, J.M., Gilbert, D.T., Axsom, D. (2000). Focalism: a source of durability bias in affective forecasting. *Journal of Personality and Social Psychology*, *78*(5), 821–36.

Young, L., Camprodon, J.A., Hauser, M., Pascual-Leone, A., Saxe, R. (2010). Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(15), 6753–8.

Young, L., Dungan, J. (2011). Where in the brain is morality? Everywhere and maybe nowhere. *Social Neuroscience*, *7*, 1–10.

Young, L., Saxe, R. (2008). An fMRI investigation of spontaneous mental state inference for moral judgment. *Journal of Cognitive Neuroscience*, *21*(7), 1396–4.

Young, L., Saxe, R. (2009). Innocent intentions: a correlation between forgiveness for accidental harm and neural activity. *Neuropsychologia*, *47*(10), 2065–72.

Zahn, R., de Oliveira-Souza, R., Bramati, I., Garrido, G., Moll, J. (2009). Subgenual cingulate activity reflects individual differences in empathic concern. *Neuroscience Letters*, *457*(2), 107–10.