Moral Judgments Recruit Domain-General Valuation Mechanisms to Integrate Representations of Probability and Magnitude

Amitai Shenhav[1,*] and Joshua D. Greene[1]
[1]Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, MA 02138, USA
*Correspondence: ashenhav@wjh.harvard.edu
DOI 10.1016/j.neuron.2010.07.020

SUMMARY

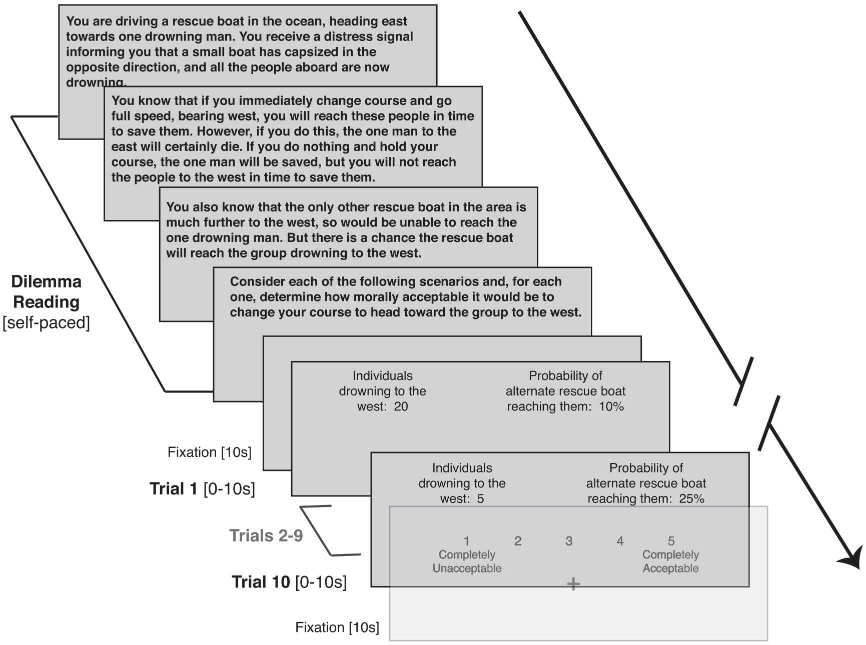
Many important moral decisions, particularly at the policy level, require the evaluation of choices involving outcomes of variable magnitude and probability. Many economic decisions involve the same problem. It is not known whether and to what extent these structurally isomorphic decisions rely on common neural mechanisms. Subjects undergoing fMRI evaluated the moral acceptability of sacrificing a single life to save a larger group of variable size and probability of dying without action. Paralleling research on economic decision making, the ventromedial prefrontal cortex and ventral striatum were specifically sensitive to the "expected moral value" of actions, i.e., the expected number of lives lost/saved. Likewise, the right anterior insula was specifically sensitive to outcome probability. Other regions tracked outcome certainty and individual differences in utilitarian tendency. The present results suggest that complex life-and-death moral decisions that affect others depend on neural circuitry adapted for more basic, self-interested decision making involving material rewards.

INTRODUCTION

The most consequential moral decisions that humans make are at the policy level, where a single choice can significantly impact thousands of lives. Examples include healthcare decisions, such as the adoption of an opt-out versus opt-in system for organ donation (Johnson and Goldstein, 2003), and military decisions, such as U.S. President Harry Truman's decision to deploy nuclear weapons against Japan. Such decisions have several notable features. First, they involve trade-offs among costs and benefits of varying *magnitude*. Second, they involve uncertainty, with outcomes that vary in their *probability* of occurrence. Third, such decisions often involve life-and-death outcomes for individuals other than the decision maker, requiring the decision maker to assess the value of these lives and incorporate such assessments into a decision. Fourth, the individuals who make policy decisions (voters, legislators, judges, government officials, etc.) are, at best, indirectly affected by the social utility of their choices and may be completely unaffected by it. The present research examines moral decisions with these four critical features, which reflect the complexity, seriousness, and indirect social nature of important policy decisions. In functional terms, the present research examines how the brain represents and integrates information concerning the magnitude and probability of outcomes in decisions with life-and-death implications for unknown others.

The present research aims to draw parallels between economic and moral decision making. This endeavor is significant in two ways. First, it addresses a central question in the study of moral judgment, namely the extent to which moral judgments draw on domain-general versus domain-specific processes (Greene and Haidt, 2002; Hauser, 2006). Some have argued that moral judgments are produced by a "moral faculty" independent of and prior to processing by affective/emotional circuitry in the brain (Hauser, 2006; Huebner et al., 2009). Evidence that such judgments are produced by domain-general, affective mechanisms of evaluation would therefore count against the hypothesis that such judgments are produced by a domain-specific moral faculty. Second, in drawing this parallel, the present research would significantly expand the purview of "neuroeconomic" models of valuation (Glimcher, 2009; Rangel et al., 2008; Wallis, 2007). Research on economic decision making has examined the neural systems responsible for tracking and integrating information concerning outcome magnitude and probability (Knutson et al., 2005; Platt and Huettel, 2008; Tom et al., 2007). However, such research has focused on decisions involving primary reinforcers or monetary outcomes for the decision maker, while the present research examines decisions involving life-and-death outcomes that affect unknown others rather than the decision maker. Thus, the present research tests the generality of neuroeconomic models that aspire to provide a comprehensive framework for subjective valuation and decision making (Glimcher, 2009; Montague and Berns, 2002). We hypothesize that the relatively detached moral decisions examined here rely on domain-general evaluation mechanisms that enable more basic, self-interested decision making in humans and animals.

Several studies have parametrically varied the *probability* and *magnitude* of positive and/or negative outcomes to identify brain regions and neurotransmitter systems responsible for representing these variables and integrating them into a subjective summary representation of expected value. Such decisions



You are driving a rescue boat in the ocean, heading east towards one drowning man. You receive a distress signal informing you that a small boat has capsized in the opposite direction, and all the people aboard are now drowning.

You know that if you immediately change course and go full speed, bearing west, you will reach these people in time to save them. However, if you do this, the one man to the east will certainly die. If you do nothing and hold your course, the one man will be saved, but you will not reach the people to the west in time to save them.

You also know that the only other rescue boat in the area is much further to the west, so would be unable to reach the one drowning man. But there is a chance the rescue boat will reach the group drowning to the west.

Consider each of the following scenarios and, for each one, determine how morally acceptable it would be to change your course to head toward the group to the west.

Dilemma
Reading
[self-paced]
Fixation [10s]
Trial 1 [0-10s]
Trials 2-9
Trial 10 [0-10s]
Fixation [10s]
Individuals drowning to the west: 20
Probability of alternate rescue boat reaching them: 10%
Individuals drowning to the west: 5
Probability of alternate rescue boat reaching them: 25%
1 Completely Unacceptable
2
3
4
5 Completely Acceptable
+

Figure 1. Task Timeline for a Single Functional Run

Subjects progress through four screens describing the dilemma context, a default action (saving one individual with certainty), and a proposed alternate action to be evaluated (saving a group of individuals who might otherwise be saved with a known probability). This is followed by ten trials in which the group size (Magnitude) and probability that the group will be saved by other means (100 – Probability) are specified. Subjects morally evaluate the proposed action using a five-point scale. Trials are self-paced (up to 10 s) and followed by a 10 s fixed intertrial interval (ITI).

involve, in different ways, subcortical regions in the striatum, thalamus, and amygdala as well as cortical regions in the cingulate cortex, insula, ventromedial prefrontal/medial orbitofrontal cortex (vmPFC/mOFC), and posterior parietal cortex (Knutson et al., 2005; Platt and Huettel, 2008; Tom et al., 2007). The vmPFC/mOFC in particular appears to be specialized for representing the overall expected value/utility associated with an option (Hare et al., 2008; Knutson et al., 2005; Wallis, 2007). We hypothesize that at least some of these neural structures will play comparable roles in the complex moral decisions examined here. Previous research on other-regarding preferences in the context of resource allocation (Hare et al., 2010; Hsu et al., 2008; Moll et al., 2006) are consistent with this hypothesis, but these studies examine decisions involving familiar economic goods and do not (explicitly) involve uncertainty. The present study, in contrast, examines representations of the value of life-and-death outcomes and how these representations are modulated by uncertainty.

The present research also builds on recent research examining hypothetical life-and-death moral dilemmas (Foot, 1978; Thomson, 1986) in which one can save several lives by sacrificing a smaller number of lives (Greene, 2009; Greene et al., 2001). Neuroscientific studies employing such dilemmas have examined several critical factors, such as the distinction between action and omission and the distinction between harm as a means and harm as a side-effect (Schaich Borg et al., 2006), but have yet to examine manipulations of probability and magnitude, which are critical for real-world, complex decision making. The present study uses a parametric design to examine these variables and their neural representation. This additionally allows us to examine individual differences in both neural and behavioral sensitivity to these variables and to examine the relationship between neural sensitivity and behavioral sensitivity to these variables.

Our first aim is to identify neural structures responsible for encoding the magnitude and probability of outcomes and subjective representations of the overall "expected moral value" of morally significant actions. Our second aim is to determine the extent to which these neural structures are consistent with those implicated in more conventional economic decision making (Knutson et al., 2005; Platt and Huettel, 2008; Sanfey et al., 2006; Wallis, 2007). More specifically, we aim to determine whether particular brain regions that are predictive of behavioral risk and value sensitivity in more conventional paradigms (Knutson et al., 2005; Paulus et al., 2003) play comparable roles in the context of moral judgment. This will be accomplished by identifying neural regions whose activity covaries with the relevant task parameters as well as regions associated with individual differences in behavioral sensitivity to these parameters. Recent studies of moral judgment (Ciaramelli et al., 2007; Greene et al., 2004, 2008; Koenigs et al., 2007) suggest that automatic emotional responses often conflict with utilitarian judgments. In light of this, our third aim is to identify neural activity associated with individual differences in willingness to endorse utilitarian trade-offs (Hsu et al., 2008) and to determine, more specifically, whether increased endorsement relies on neural circuitry involved in emotion regulation (Hooker and Knight, 2006; Wager et al., 2008). Our final aim is to identify patterns of neural activity associated with decisions in which the outcomes are more versus less certain, testing the hypothesis that moral judgments involving more certain outcomes are more likely to depend on mechanisms for rule-guided choice, typically enabled by the lateral PFC (Badre and D'Esposito, 2007; Greene et al., 2004; Miller and Cohen, 2001).

RESULTS

Subjects undergoing fMRI judged the moral acceptability of actions (e.g., turning a rescue boat away from a drowning man) that would result in the certain death of one individual (Figure 1). Each such action would also prevent (with certainty) the deaths of a group of individuals (e.g., a group drowning in the opposite direction). These deaths would otherwise be prevented with

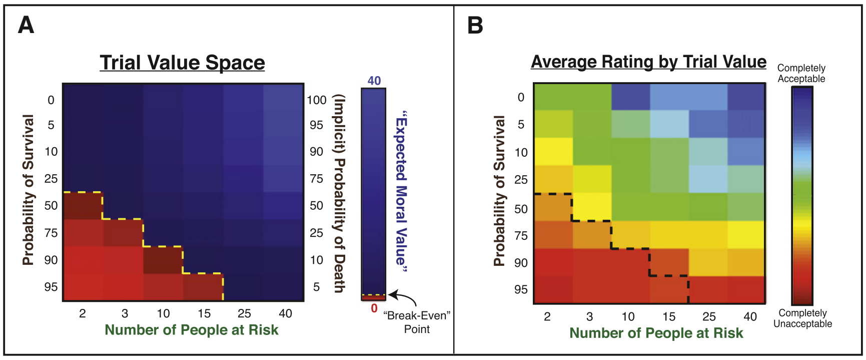


Figure 2. Distribution of Trial Values and Ratings of Moral Acceptability

Subjects' judgments did not reverse sharply when the relative values of the two options reversed. Instead they exhibited a more graded sensitivity to "Expected Moral Value."

(A) The trial value space from which Magnitude-Probability pairs were randomly drawn for each trial. Forty-eight possible trial/action types are color coded by Expected Value (number of lives expected to be saved through action). Actions/trials with an Expected Value < 1 (the amount required to match the certain death of one individual abandoned by acting) are in red, and those with value > 1 (net gain) are in blue.

(B) Average moral acceptability ratings across trial value space reveal a graded behavioral sensitivity to "Expected Moral Value" and its components.

See also Figure S1.

some known probability (e.g., by an alternate rescue boat). This probability and the outcome magnitude (size of group saved) were varied parametrically across trials, thus varying the overall expected value of performing the action, i.e., the expected number of lives saved/lost. Figure 2A shows the entire set of possible trial values, color-coded according to the expected value of the proposed action (defined here as a simple multiplicative interaction of outcome magnitude and probability). Blue and red, respectively, indicate values above and below the "break-even point," i.e., the set of points at which the number of expected lives saved and the number of expected lives lost both equal 1.

Behavioral Results

Subjects' ratings of the moral acceptability of proposed actions were examined using a mixed-effects multiple regression model with subject entered as a random effect. Judgments were highly sensitive to the number of lives at stake (Magnitude, natural log-transformed) (t(33,1) = 9.38), the probability that these lives would be lost without action (Probability) (t(33,1) = 15.1), and the Expected Value of the action (Magnitude × Probability) (t(33,1) = 4.06) (All p values < 0.0005; See Figure 2B). The directions of these effects were consistent with classical economic models, with subjects' judging it more acceptable to actively sacrifice a life when inaction involves larger and more probable losses of life. Contrary to models based on linearly increasing utility from lives saved, subjects' indifference thresholds (based on acceptability ratings) were shifted upward from the break-even point and did not drop off sharply at any point (see Figure 2B). A follow-up experiment verified that this shift in indifference from the break-even point was not a consequence of using a Likert scale rather than a binary choice between outcomes (see Experimental Procedures and see Figure S1 available online). A second follow-up experiment verified that a significant shift in indifference away from the break-even point is still present when a Loss frame (lives will be *lost* without action) is used instead of a Gain frame (lives will be *saved* by acting). (one-sample $t(21) = 2.47$, two-tailed $p < 0.05$; see Figure S1D) Consistent with previous research (Petrinovich and O'Neill, 1996; Tversky and Kahneman, 1981), subjects were marginally less risk averse (i.e., more willing to take a chance on saving the larger group) under the Loss frame (comparison of regression intercepts: two-sample t(43) = 1.64, one-tailed p = 0.055; Figure S1). Likewise, contrary to normative models that value all lives equally, acceptability ratings more closely tracked a natural-log transformation of magnitude ($R^2 = 0.60$) than a linear function of magnitude ($R^2 = 0.48$). This is consistent with the implication of a primitive, analog system for representing approximate magnitudes using a logarithmic scale (Dehaene et al., 1999; Nieder and Miller, 2003), as well as psychophysical models of sensory processes (Thurstone, 1954) and models of economic valuation and diminishing marginal utility (Bernoulli, 1954). (Henceforth, we use "Magnitude" to refer to ln(Magnitude).) RT's were not significantly influenced by Magnitude and Probability (t(32.3,1) = −1.37, p = 0.18; t(32.9,1) = 1.23, p = 0.23, respectively), but were faster as Expected Value increased (t(32.1,1) = −3.75, p = 0.0007).

Neuroimaging Results

Magnitude, Probability, and Expected Value

We performed whole-brain analyses to identify brain regions exhibiting parametric increases in BOLD signal specifically tracking increases in outcome Magnitude, Probability, and Expected Value (Magnitude × Probability interaction). We observed a positive correlation with Magnitude in bilateral regions of anterior and posterior cingulate cortices, central insula, putamen, and inferior parietal lobe, among other regions (Figure 3A; Table 1). We found comparable sensitivity to Probability (greater signal in response to higher probability of loss through inaction, raising the relative value of action) in left dorsal posterior insula, putamen, ventral posterolateral thalamus, and right posterior cingulate cortex, among other regions (Figure 3B; Table 1). No regions exhibited significant effects opposite to these. Our serially orthogonalized regression procedure allowed us to examine parametric sensitivity to Expected Value over and above sensitivity to its two components (Magnitude and Probability). We observed positive correlations between BOLD signal and Expected Value bilaterally in vmPFC/mOFC, precuneus, and inferior parietal lobe, left mediodorsal thalamus, left ventrolateral PFC, and left superior temporal sulcus (Figure 4; Table 1). No opposite effects were observed.

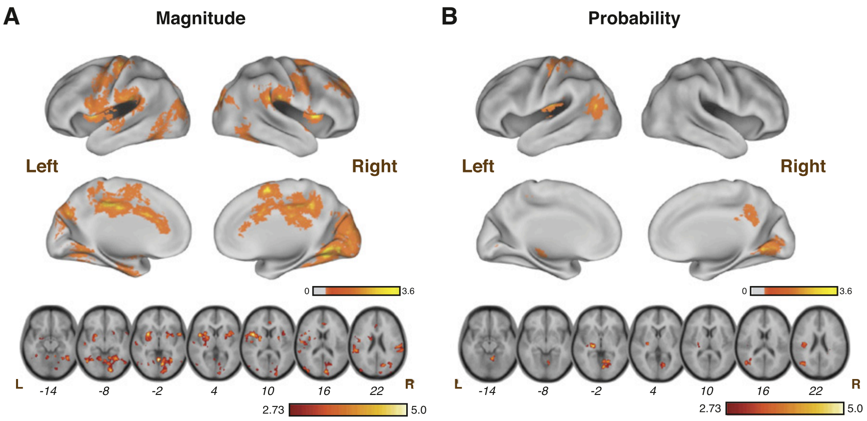


Figure 3. Brain Regions Tracking Variations in Magnitude and Probability

(A) Regions including the posterior cingulate cortex, central insula, putamen, and inferior parietal lobe exhibited significant parametric increases in BOLD activity with logarithmically increasing number of lives potentially saved/lost (Magnitude).

(B) Regions including the left dorsal posterior insula, putamen, ventral posterolateral thalamus, and right posterior cingulate cortex exhibited significant parametric increases in BOLD activity with linearly increasing probability that the action is necessary to save the group (Probability). No regions exhibited effects opposite those shown in (A) and (B). Statistical maps represent t values thresholded at a voxelwise threshold of $p < 0.005$ and a clusterwise threshold of $p < 0.05$.

Behavioral and Neural Sensitivity to Probability of Loss of Life

The above whole-brain analyses may fail to identify effects that are modulated by individual differences in sensitivity to our parameters of interest. For this reason we performed additional analyses to interrogate a priori ROIs previously identified as relevant to individual differences in analogous decision tasks. BOLD activity in the right anterior insula (R-aIns) has been shown to increase with the contemplation of riskier choices, and to correlate with avoidant traits both in individual studies (Kuhnen and Knutson, 2005; Paulus et al., 2003; Venkatraman et al., 2009) and in a recent meta-analysis of the literature on risk processing (Mohr et al., 2010). This suggests that BOLD signal in this region may correlate with individual differences in sensitivity to the Probability parameter (i.e., increased probability of the group's death) in the current paradigm. To test this hypothesis, we extracted averaged Probability contrast estimates (i.e., BOLD sensitivity to Probability) for each subject from an a priori spherical ROI centered on the R-aIns focus of activation in Paulus et al. (2003) (Figure 5A). We compared these individual estimates of R-aIns sensitivity to an estimate of how sensitive each subject's moral judgments were to the Probability parameter (i.e., least-squares estimates from a regression of acceptability ratings on each of our parameters). As predicted, we found that neural sensitivity to Probability in the R-aIns was correlated with behavioral sensitivity to Probability ($r = 0.38$, $p = 0.02$; Figure 5B). That is, the degree to which an individual's moral judgments were influenced by the likelihood of the group's death was predicted by the degree to which that individual's R-aIns was sensitive to this same parameter.

Behavioral and Neural Sensitivity to Expected Value

Regions of ventral striatum (vStr) have been consistently shown to encode reward value and reward prediction errors across a wide array of reinforcers (Knutson et al., 2005; O'Doherty et al., 2004). Likewise, activity in these regions is known to correlate with individual sensitivity to reward gains relative to losses (Tom et al., 2007; Venkatraman et al., 2009). We extracted averaged Expected Value contrast estimates (i.e., BOLD sensitivity to Expected Value) for each subject from bilateral ROI's centered on the vStr foci of activation in Knutson et al. (2005) (Figure 5C). As predicted, these estimates were significantly correlated (left: $r = 0.51$, $p = 0.005$; right: $r = 0.55$, $p = 0.002$) with individual behavioral Expected Value beta estimates (i.e., sensitivity of moral judgments to Expected Value) (Figure 5D). Thus, across individuals, BOLD sensitivity to "expected moral value" (Magnitude × Probability) in the ventral striatum correlates with behavioral sensitivity to that factor.

Individual Differences in Utilitarian Tendency

Individuals vary in their tendency toward utilitarian judgment (Bartels, 2008; Greene et al., 2004), which in the present context refers to the approval of allowing one person's death in order to save a larger number of lives. Across subjects, higher mean acceptability ratings reflect greater approval of utilitarian tradeoffs. We performed a whole-brain analysis across subjects, regressing each subject's average BOLD signal during the trial period (relative to fixation baseline) against his/her average acceptability rating. We found that increased utilitarian responding at the individual level was correlated with increased BOLD activity in bilateral regions of lateral OFC/vlPFC and medial superior frontal gyrus, left middle temporal gyrus and superior parietal lobe (Figure 6; Table S1). No regions exhibited the opposite pattern of activity.

Encoding/Response to Outcome Certainty

It is likely that moral judgments made under conditions of uncertainty are different from those made under conditions of certainty. Previous research employing moral dilemmas with certain outcomes has associated utilitarian judgment with activity in the dorsolateral prefrontal cortex (dlPFC) and corresponding regions in the parietal lobes, and it has been proposed that such effects reflect the application of a utilitarian decision rule (Greene et al., 2001, 2004). This suggests that such regions will exhibit greater activity in the present context as the outcomes in question become more certain. To identify brain regions sensitive to outcome certainty, we included in our full parametric regression model (along with ln(Magnitude), Probability, and Expected Value) a quadratic Probability term describing a "U-shaped" response profile for BOLD activity across probabilities (i.e., activity varying with increasing distance from 50% chance of group survival). This analysis revealed regions exhibiting increased BOLD signal with increasing certainty in bilateral vlPFC,

Table 1. Result of a Whole-Brain Analysis Identifying Brain Regions Sensitive to Outcome Magnitude, Probability, and Expected Value

Regressor	Side	Region(s)	Cluster Extent	Peak Z score	Peak MNI Coordinates		
			(voxels)		X	Y	Z
Magnitude							
	L	cIns, putamen, premotor	2211	4.68	−42	4	6
	B	Cerebellum, lingual gyrus, R extrastriate	3819	4.55	−2	−60	−4
	R	Premotor, precentral gyrus	268	4.49	52	−2	40
	B	ACC, SMA	925	4.30	8	6	62
	L	Inferior temporal cortex, fusiform gyrus	208	4.02	−46	−62	−10
	B	PCC	1189	4.01	8	−30	48
	L	Anterior parahippocampal gyrus	250	3.79	−30	−12	−30
	R	cIns, putamen	520	3.79	54	8	−2
	L	IPL, SII	363	3.76	−56	−34	24
	R	SFG, MFG	279	3.73	26	48	36
	R	IPL, SII	448	3.71	64	−36	26
	L	LOC	216	3.68	−42	−86	18
Probability							
	R	Lingual gyrus, cerebellum	530	4.17	6	−74	0
	L	VPL thalamus, putamen	198	4.15	−20	−22	−2
	L	LOC	243	3.99	−46	−66	20
	L	Dorsal posterior insula	215	3.93	−38	−22	26
	L	Precentral/postcentral gyri	299	3.69	−16	−28	62
	R	PCC	239	3.56	28	−36	36
Expected Value							
	L	STS	735	4.17	−46	−10	−22
	B	PCC, precuneus, RSC	1450	4.03	8	−64	24
	B	vmPFC/mOFC, frontal pole	758	3.90	−4	56	−12
	L	MD/VL thalamus	243	3.89	−14	−14	6
	L	IPL	452	3.85	−42	−60	18
	L	vlPFC, IFG	186	3.80	−52	30	0
	R	IPL	229	3.63	52	−54	26

Regions exhibiting significant parametric increases in BOLD activity with linearly increasing (A) Magnitude (larger number of lives saved through action), (B) Probability of saving lives through action, (C) Expected Value of proposed action. Significant clusters met a voxelwise threshold of $p < 0.005$ and a clusterwise threshold of $p < 0.05$. L, left; R, right; B, bilateral. cIns, central insula; ACC, anterior cingulate cortex, SMA, supplementary motor area; SII, secondary somatosensory area; MFG, middle frontal gyrus; SFG, superior frontal gyrus; LOC, lateral occipital cortex; PCC, posterior cingulate cortex; VPL, vental posterolateral; STS, superior temporal sulcus; IPL, inferior parietal lobule; RSC, retrosplenial cortex; vmPFC, ventromedial prefrontal cortex; mOFC, medial orbitofrontal cortex; vlPFC, ventrolateral prefrontal cortex; MD, medial dorsal; VL, ventral lateral; IFG, inferior frontal gyrus.

precuneus, inferior parietal lobe, SMA, SFG, and left lateral PFC (Figure 7; Table S2). No regions exhibited significant reverse effects.

DISCUSSION

The present study examined the neural mechanisms responsible for making complex moral decisions involving outcomes (lives saved/lost) of variable magnitude and probability and hence options that varied in expected value. Our results indicate that the mechanisms that enable such decisions overlap considerably with those that enable more familiar self-interested decisions (Knutson et al., 2005; Platt and Huettel, 2008; Tom et al., 2007). Most notably, we found that BOLD signal in the vmPFC/mOFC correlated with the “expected moral value” of decision options, i.e., the interaction between magnitude and probability. This is consistent with the hypothesis that this region supports the integration of positive and negative reward signals into a more abstract representation of value, a kind of decision “currency” (Chib et al., 2009; Hare et al., 2008; Kringelbach and Rolls, 2004; O’Doherty et al., 2001; Padoa-Schioppa, 2007; Wallis, 2007). The present results suggest that the vmPFC/mOFC, in addition to representing the subjective value of material gains and losses that may accrue to the decision-maker, has been recruited to represent hypothetical gains and losses, including gains and losses of life, that may accrue to others, but that have no material bearing on the decision maker. While we have used an objective stimulus parameter (expected

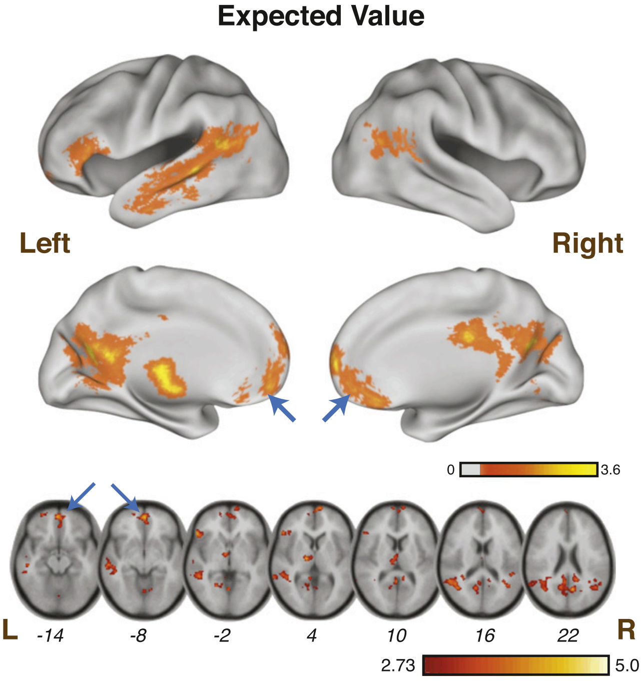


Figure 4. Regions Exhibiting Significant Parametric Increases in BOLD Activity with Linearly Increasing Number of Lives Expected to Be Saved (Expected Value)

Paralleling results of several studies of economic decision making, the ventromedial prefrontal cortex (see arrows) was specifically sensitive to the "expected moral value" of actions. Other regions, including the precuneus, inferior parietal lobe, left mediodorsal thalamus, left ventrolateral PFC, and left superior temporal sulcus, exhibited this effect as well. No regions exhibited opposite effects. Statistical maps represent t values thresholded at a voxelwise threshold of $p < 0.005$ and a clusterwise threshold of $p < 0.05$.

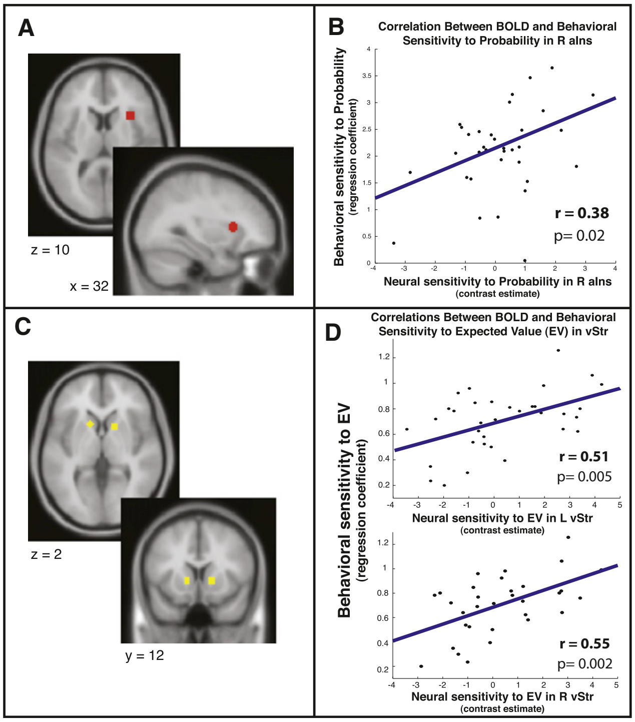


Figure 5. Correlations between Behavioral and Neural Sensitivity to Risk and Value within Three Spherical A Priori ROIs

(A) A region of right anterior insula identified by Paulus et al. (2003) as sensitive to risk and correlated with harm avoidance.

(B) Individuals' neural sensitivity to Probability in this region of insula correlates with individuals' behavioral sensitivity to Probability.

(C) Bilateral regions of ventral striatum identified by Knutson et al. (2005) as sensitive to reward value.

(D) Individuals' neural sensitivity to expected number of lives saved (Expected Value) in these regions of ventral striatum correlates with individuals' behavioral sensitivity to Expected Value. Pearson's r values are supplemented by p values from a robust regression.

value) to identify activity in this region, it is unlikely that the value representations it encodes correspond precisely to this objective parameter. Rather, the activity in this region more likely reflects a subjective valuation function that is sensitive to expected value, but also sensitive to factors related to the individual's prior experience and present context. This is consistent with our finding that subjects' judgments are sensitive to gain/loss framing and appear to be better predicted by a log-transformed magnitude function. Our whole-brain search for brain regions sensitive to expected moral value and its components revealed effects in a number of regions in addition to the vmPFC/mOFC (see Table 1). The respective contributions of these brain regions to complex moral decision making remains a topic for future research.

Also consistent with studies of economic decision making (Knutson et al., 2005; Tom et al., 2007), we found that BOLD signal in regions of central insula, dorsal striatum, and anterior and posterior cingulate cortices was sensitive to the magnitude of the potential loss/gain associated with decision options. Also consistent with previous studies of decision making (particularly ones involving the assessment of risk) (Paulus et al., 2003; Qin and Han, 2009), we found that BOLD signal in the left posterior insula was sensitive to the probability of loss/gain associated with decision options. Among the key limitations of the present paradigm is that valence at encoding cannot be determined based on the task alone. That is, a larger group to be rescued may be seen as positive (more people who can be saved) or negative (more people in danger). The pattern of activations in the ventral posterolateral thalamus, cingulate cortex, secondary somatosensory cortex, and insula for the Magnitude and/or Probability regressors suggests aversive representations similar to those observed in anticipation of, and response to, nociceptive stimuli (Craig, 2002; Price, 2000) and to more abstract negative outcomes (Bechara and Damasio, 2005; Knutson and Greer, 2008). We note, however, that these results alone are insufficient to make strong inferences concerning the valence of the representations of these parameters. We conducted a follow-up experiment to determine whether subjects viewed the available outcomes as consistently negative, consistently positive, or not consistently either. Self-report data were collected following the Likert scale and binary choice experiments described above (both gain frame; see Supplemental Information). A majority of participants, 54.4%, reported experiencing the dilemmas as

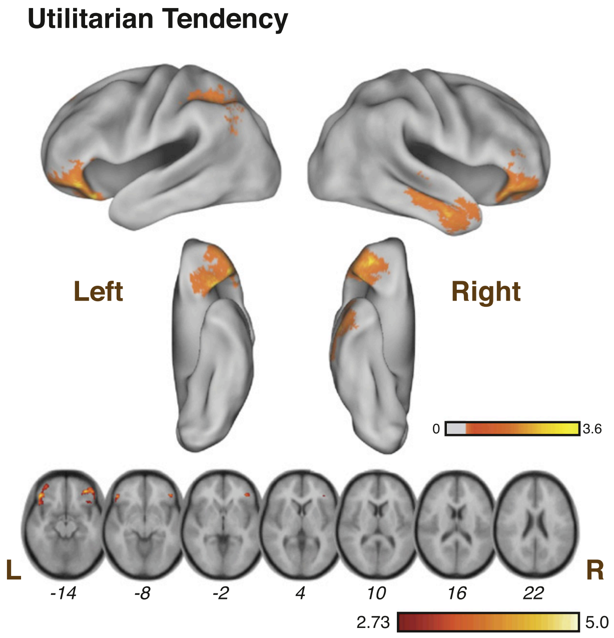


Figure 6. Regions Exhibiting Significant Parametric Increases in BOLD Activity with Increased Tendency toward Utilitarian Judgment at the Individual Level

These include bilateral regions of lateral OFC/vlPFC, medial superior frontal gyrus, left middle temporal gyrus, and superior parietal lobe. No regions exhibited opposite effects. Statistical maps represent t values thresholded at a voxelwise threshold of $p < 0.005$ and a clusterwise threshold of $p < 0.05$. See also Table S1.

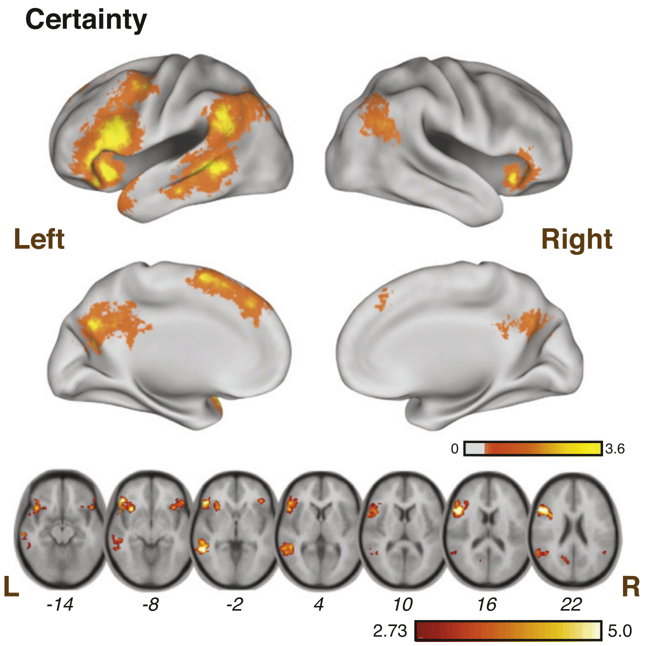


Figure 7. Regions Exhibiting Significant Parametric Increases in BOLD Activity with Increased Certainty of Saving/Losing Lives If No Action Is Taken (Quadratic Function of Probability)

These include bilateral vlPFC, precuneus, inferior parietal lobe, SMA, SFG, and left lateral PFC. No regions exhibited opposite effects. Statistical maps represent t values thresholded at a voxelwise threshold of $p < 0.005$ and a clusterwise threshold of $p < 0.05$. See also Table S2.

decisions between two bad outcomes. 17.6% reported experiencing them as between two good outcomes, and 28% viewed the outcomes as having mixed valences.

We uncovered further parallels between complex moral decision making and economic decision making at the level of individual differences. Two brain regions, the right anterior insula and the ventral striatum, were selected a priori for analysis based on prior findings identifying them, respectively, as sensitive to risk (Mohr et al., 2010; Paulus et al., 2003; Venkatraman et al., 2009) and reward value (Knutson et al., 2005; Tom et al., 2007; Venkatraman et al., 2009; Yacubian et al., 2007). As predicted, we found that individual differences in behavioral sensitivity to the probability of others' deaths (the parameter most closely approximating "risk," as defined in the relevant literature) were correlated with individual differences in BOLD sensitivity to this parameter in the right insula. Comparable effects were observed bilaterally in the ventral striatum, correlating neural sensitivity to expected value with behavioral sensitivity to expected value. Thus, our results suggest that an individual's sensitivity to lives saved/lost in the context of moral judgment is in part determined by the same mechanisms that determine that individual's sensitivity to the probability of loss and to overall reward in the context of self-interested economic decision making.

Subjects exhibited systematically different patterns of neural activity in response to moral decisions involving more versus less uncertainty. Trials involving more certain outcomes (i.e., nearer to 0% or 100% outcome probability) elicited increased activity in ventrolateral and left lateral regions of PFC. This may be because these trials require less weighing and integration across parameters (Magnitude and Probability) and may instead be made simply on the basis of Magnitude. This interpretation is consistent with research on humans and nonhuman primates implicating the lateral PFC (especially left hemisphere) in "top down" action selection in the service of achieving more distant/abstract goals (Badre and D'Esposito, 2007; Koechlin et al., 2003; Miller and Cohen, 2001).

Little is known about the neural bases of individual differences in utilitarian tendency, i.e., the tendency to favor options maximizing aggregate welfare over options with other desirable features, such as promoting equality (Hsu et al., 2008). We found that increased activity in bilateral lateral OFC (anterior and ventral to the OFC region identified in association with outcome certainty) was associated with more frequent endorsement of utilitarian trade-offs. According to Greene et al.'s dual-process theory of moral judgment (Greene et al., 2001, 2004, 2008), utilitarian judgments are driven primarily by controlled cognitive processes, which may compete with countervailing emotional responses. These results are broadly consistent with this dual-process theory, given the implication of lateral OFC in reducing the influence of emotional distracters on judgments (Beer et al., 2006b), regulating pain and negative emotions (Phan et al., 2005; Wager et al., 2004), favoring delayed rewards over more immediate ones (Boettiger et al., 2007), inhibiting socially inappropriate behaviors (Beer et al., 2006a), and more generally controlling the influence of emotional responses that interfere with the pursuit of more distal goals (Beer et al., 2006b; Hooker and Knight, 2006; Kim and Hamann, 2007; Watanabe and Sakagami, 2007). However, previous research has associated utilitarian judgment with increased activity in the dorsolateral PFC (Greene et al., 2004), rather than lateral OFC. These two findings may perhaps be reconciled as reflections of different types of affective regulation, corresponding to the demands of two related, but functionally distinct, moral judgment tasks. It was hypothesized that previous results (Greene et al., 2004) reflect the regulation of emotions through effortful cognitive control whereas, in contrast, the present results may reflect more implicit modulation of affective representations. In other words, the present use of repeated, incrementally varied affective stimuli may elicit a modulation of ‘‘bottom-up’’ affective influences on judgments, consistent with the aforementioned literature implicating the lateral OFC in performing a gating or weighing function (Beer et al., 2006b; Rule et al., 2002; Wager et al., 2004), as opposed to the overriding of a prepotent response (Cunningham et al., 2004). We also note that the present effect in the OFC reflects individual differences, as opposed to within-individual differences between response types. An alternative account of the effect of utilitarian tendency in the OFC derives from an alternative account of OFC function according to which the value of positively valenced outcomes are encoded more medially and negatively valenced outcomes more laterally (Kringelbach and Rolls, 2004; Ursu et al., 2008). Thus, it could be that subjects who rated saving the group as more acceptable were guided by more robust negative representations of the deaths of those individuals.

Our general finding of considerable overlap between economic decision making and moral decision making (of the present sort) is notable for several reasons. As noted above, the choices our subjects faced concerned saving lives as opposed to more modest rewards/punishments, concerned outcomes for others rather than for the decision maker him/herself, were hypothetical (by ethical necessity), and required the generation of ratings of moral acceptability rather than the indication of a personal preference. In light of these features, one might expect the present judgments to depend primarily on neural circuitry that is independent of those responsible for the representation of value in the context of basic, self-interested decision making, and yet that was not the case. We emphasize that our use of hypothetical decisions (ordinarily considered a limitation) underscores the present conclusions: even moral decisions concerning outcomes for hypothetical others depend on mechanisms engaged by basic, self-interested decision making with real material rewards. We also note that our findings are broadly consistent with recent research on the neural bases of decisions to make charitable donations (Hare et al., 2010; Hsu et al., 2008; Moll et al., 2006). The convergence between our results and those involving more familiar rewards (e.g., food, money) speaks to the flexibility and domain generality of valuation circuitry highlighted here (Rangel et al., 2008). Likewise, in implicating domain-general mechanisms, our results speak against the hypothesis that moral judgments are produced by a dedicated, domain-specific ‘‘organ’’ for moral judgment (Hauser, 2006; Huebner et al., 2009). With respect to this issue, we note that the processes implicated here are not implicated as secondary modulatory processes, but rather as core evaluative processes, much as they are implicated elsewhere.

More generally, these findings may illuminate the mechanisms behind some of our most important social decisions, namely policy decisions involving uncertainty and life-and-death stakes for large numbers of people. Research on judgment and decision making indicates that such judgments often rely on heuristics and are, for this reason among others, subject to systematic biases (Kahneman, 2003; Slovic, 2007). Recent research has examined the neural bases of such biases, implicating many of the structures identified in the present research (De Martino et al., 2006, 2009; Tom et al., 2007). If, as the present results suggest, the neural mechanisms we use to think about complex, life-and-death moral decisions are in fact mechanisms originally adapted primarily for other purposes (e.g., foraging for food), then it becomes more likely that such decisions are made suboptimally. While the present results do not underwrite any specific policy recommendations, it is possible that a better understanding of our most basic and general decision-making systems will fruitfully illuminate the strengths and limitations of the capacities upon which we rely in making socially significant moral decisions.

EXPERIMENTAL PROCEDURES

Subjects

Forty right-handed subjects (twenty female) with no reported history of neurological/psychological disorder were recruited for this study. Of these, five were excluded prior to fMRI analysis—one for an incomplete session, one for a neurological abnormality that was discovered during the imaging session, one for excessive response failure (30%), and two for failing our catch trial criteria (see below). One additional subject was excluded due to excessive MR signal artifact. Following exclusions, data from 34 subjects (17 female, mean age 24.3, age range 18–42 years old) were analyzed.

Imaging Methods

Images were acquired using a 3.0 T Siemens Magnetom Tim Trio full-body scanner at the Martinos Center for Biomedical Imaging of Massachusetts General Hospital. A high-resolution, whole-brain structural scan (1 mm isotropic voxel MPRAGE) was acquired prior to functional imaging. T2*-weighted functional images were acquired in 36 axial slices parallel to the AC-PC line with a 0.5 mm interslice gap, affording full-brain coverage. Images were acquired using an EPI pulse sequence, with a TR of 2500 ms, a TE of 28 ms, a flip angle of 90, an FOV of 256 mm and 96 × 96 matrix (resulting in 3.0 mm isotropic voxels). Four additional images included at the start of each functional run to allow for signal stabilization were discarded. Stimulus presentation and response collection were performed using Psychtoolbox (http://www.psychtoolbox.org) running on Matlab (http://www.mathworks.com).

fMRI Task

Subjects were presented with 5 different scenarios (one per run) in which they evaluated the moral acceptability of actions within the context of that scenario (Figure 1). In all scenarios a proposed action (e.g., redirecting a rescue boat) would result in the certain death of one individual but would save the lives of a group of other individuals with a specified probability. The initial scenario descriptions left unstated the group size (Magnitude) and likelihood that the group will survive if the proposed action is not taken (Probability).

After reading each scenario description, subjects completed ten trials in which variable values of Magnitude and Probability were provided (on L and R screen, respectively) with a short description of what the values represent within the scenario. Subjects were given up to ten seconds to respond by pressing one of five buttons (right hand), rating the moral acceptability of the action on a five-point scale ranging from "Completely Unacceptable" (1) to "Completely Acceptable" (5). This scale was used in place of a dichotomous forced-choice response to reduce the likelihood that subjects would use an explicit strategy (e.g., based on threshold values of Magnitude and/or Probability). A 10 s fixation ITI followed each response as well as the initial scenario description. An additional 5 s of fixation concluded each run. In total, subjects completed five fMRI runs, each consisting of 10 trials, for a total of 50 trials.

Magnitude and Probability values for each trial were chosen pseudorandomly from 50 unique pairs derived from a full factorial model containing six levels of Magnitude (2, 3, 10, 15, 25, 40 lives) and eight levels of Probability (0%, 5%, 10%, 25%, 50%, 75%, 90%, 95% likelihood of alternative rescue). The "expected moral value" (Expected Value) of a given trial/action is the product of the Magnitude and Probability parameters (Figure 2A). Trials in which the Expected Value of action exceeds the Expected Value of inaction are shown in blue. Subjects completed two sets of four practice trials (two separate scenarios) prior to entering the scanner. We included two additional catch trials with Probability set to 100% and Magnitude selected randomly from among the six levels. Subjects were excluded if they rated they rated the actions in both catch trials higher than 2 (i.e., endorsing the abandonment of a victim in order to save others who will be saved no matter what). Catch trials and trials in which the subject failed to respond within 10 s were modeled together as a condition of no interest within the fMRI analysis (see below). See Supplemental Information for complete testing materials.

Image Processing and Analysis

Imaging analysis was performed using SPM2 (Wellcome Department of Imaging Neuroscience, Institute of Neurology, London, UK). Images were spatially aligned to the first volume, then spatially normalized to a standard T2* template and resampled to 2 mm isotropic voxels, then smoothed using a Gaussian kernel (FWHM 6 mm). Each run was individually high-pass filtered using a cut-off period of 128 s.

Individual subject data were analyzed using a general linear model (GLM). Events were modeled with a variable-duration boxcar function convolved with a canonical hemodynamic response function (HRF), including an additional regressor modeling the first temporal derivative. Within each run, the initial reading event (reading the general scenario description) as well as any error trials were modeled as separate event types that were also convolved with canonical HRF's as regressors of no interest. Judgment events (moral evaluation of proposed actions given specific Magnitude and Probability values) were modeled beginning 1 s following the onset of the prompt to allow for reading/encoding of the Magnitude/Probability values. Judgment trials were modeled with parametric regressors modeling (in the following order) reaction time (RT), Magnitude (the number of people potentially saved by the action, natural-log transformed), Probability (the probability that the group will die if the proposed action is not taken, i.e., 1 minus the probability of alternative rescue, as shown to subjects), and Expected Value (the number of lives expected to be saved, i.e., the multiplicative interaction of Magnitude and Probability). All effects were modeled as linear functions of parameters with the exception of the quadratic function used to model the effect of Probability in the analysis of (un)certainty. (High certainty corresponds to both high and low Probability values.) Regressors were serially orthogonalized, with later-entered regressors accounting only for variance unaccounted for by earlier-entered regressors. Thus, all effects are controlled for RT and effects of Expected Value are effects over and above its individual components.

First-level (single subject) contrasts were performed separately over the mean judgment trial activity (relative to baseline) and for each of the parametric regressors described above (with the exception of RT). Second-level (group) random effects analyses then proceeded by performing one-sample t tests over each of these contrasts. For whole-brain analyses, resulting statistical maps were submitted to a voxelwise threshold of $p < 0.005$ and a cluster extent threshold resulting in a whole-brain corrected cluster-wise $p < 0.05$. Cluster extent thresholds were determined separately for each analysis using John Ashburner's CorrClustTh script (http://www.sph.umich.edu/~nichols/JohnsGems2.html), resulting in minimum thresholds ranging from 178 to 206 voxels in extent.

Planned region-of-interest (ROI) analyses to explore behavioral/neural sensitivity to risk and value were performed in right anterior insula and bilateral ventral striatum, respectively. Spherical ROI's were centered on MNI-coordinate space transformed peak activations from Paulus et al. (2003) (MNI x,y,z: 32,18,9; 6 mm radius) and Knutson et al. (2005) (MNI x,y,z: −15,11,1; 14,14,1; both 4.5 mm radius). Standardized contrast estimates for the relevant parametric regressor were extracted and averaged across each anatomical ROI. Behavioral sensitivity for Probability and Expected Values were calculated as regression coefficients derived from independent single-subject regressions of acceptability ratings on those single regressors. These measures of behavioral sensitivity thus reflected the degree to which the given parameter (Probability or [log-]Expected Value) influenced a subject's judgment of allowing an individual to die. The contrast estimates extracted from ROIs (R-aIns or vStr, respectively) likewise reflected the degree to which a subject's BOLD signal in that region was influenced by the relevant parameter. These two estimates (behavioral and neural) were correlated to test whether the neural sensitivity of a given region (e.g., vStr) to a specific parameter (e.g., Expected Value) predicted the sensitivity of their judgments to this same parameter (i.e., the degree to which their ratings of acceptability were influenced by changes in this parameter). In order to minimize any influence of outliers, tests of significant correlations between neural and behavioral sensitivity estimates were calculated using robust regression (iteratively reweighted least-squares approach), supplementing the Pearson's *r* values reported.

Visualization of Results

CARET software (http://brainmap.wustl.edu) was used to map group-level statistical maps (volumetric maps thresholded and masked to include only clusters that met the criteria described above) onto a cortical surface rendering, using the Probablistic Average Landmark and Surface-Based (PALS) atlas (Van Essen, 2005). A multifiducial mapping technique (Van Essen, 2005) was employed and surface statistical values were interpolated, with the goal of achieving better estimates of cluster extent along the cortical surface at the cost of precision in absolute statistical values. For completeness, surface renderings are shown alongside axial slices of an averaged MNI structural volume showing the group statistical maps which are being projected. Slices were chosen to focus more directly on subcortical structures.

Follow-Up Behavioral Experiments

Two additional experiments were performed outside of the MR scanner to examine the generalizability of the behavioral results of our fMRI experiment. All task parameters, testing materials, and exclusionary criteria were identical to those used in the original task unless otherwise stated.

Choice Task

Subjects (n = 22, 13 female) completed a task identical to the fMRI task except that participants were instructed to provide responses given a binary choice rather than a graded moral acceptability scale. Rather than responding using a 1–5 scale appearing at the bottom of each trial slide, subjects indicated whether they preferred to "Stay" or "Switch" (mapped to key presses 1 and 2, respectively). The final slide of the dilemma text before they began the trials reminded them which option corresponded to Stay (e.g., continue on course to save one drowning man) versus Switch (e.g., change course to save the group instead). Choices were regressed against log-Expected Value, using a logit model, and each subject's resulting model was used to determine the expected value of lives saved at each subject's indifference point, i.e., the point at which the subject is equally likely to Stay or Switch. These expected value estimates were restricted to the range of expected values presented to subjects (0.1–40), such that values less than ln(0.1) or greater than ln(40) were set to ln(0.1) or ln(40), respectively. This limited the impact of extreme values and thus made for a more conservative test of our hypotheses, given our prediction that mean EV at the indifference point will be greater than 1.

Framing Task

Two groups of subjects completed a task identical to the fMRI task except that, for one group, the descriptions of the options employed a Loss frame rather than a Gain frame. The Gain frame group (n = 23, 12 female) was presented with options described in terms of the number of lives that could otherwise be *saved*, while the Loss frame group (n = 21, 10 female) was presented with materially identical options described in terms of the number of lives that could be *lost*. (see Supplemental Information for wording details). For the Loss frame group, probability values were presented as the original Probability values subtracted from 100. Ratings were regressed (using robust regression) against Expected Value (natural-log transformed) and each subject's resulting model was used to determine the expected value of lives saved at each subject's indifference point (the midpoint of the 1–5 scale). In order to directly compare the acceptability ratings and binary choice data, an additional analysis was performed whereby ratings from the Gain frame group were transformed to binary choices (ratings of 1–2 as ''Stay,'' 4–5 as ''Switch,'' and trials with a response of 3 excluded). The expected value at each subject's indifference point was determined as described above for the choice task data.

SUPPLEMENTAL INFORMATION

Supplemental Information includes one figure, two tables, and Supplemental Experimental Procedures and can be found with this article online at doi:10.1016/j.neuron.2010.07.020.

ACKNOWLEDGMENTS

We are grateful to R. Bernhard, K. Hermann, D. Murray, J. Paxton, and P. Sanchez for assistance in data collection, R. Buckner and Y.C. Kao for technical guidance, and A. Bornstein, F. Krienen, and three anonymous reviewers for their thoughtful comments on the manuscript. This work was supported by an NSF Graduate Research Fellowship awarded to A.S. and NSF SES-082197 8 awarded to J.D.G.

Accepted: July 20, 2010
Published: August 25, 2010

REFERENCES

Badre, D., and D'Esposito, M. (2007). Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. J. Cogn. Neurosci. *19*, 2082–2099.

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

Bechara, A., and Damasio, A. (2005). The somatic marker hypothesis: A neural theory of economic decision. Games Econ. Behav. *52*, 336–372.

Beer, J.S., John, O.P., Scabini, D., and Knight, R.T. (2006a). Orbitofrontal cortex and social behavior: integrating self-monitoring and emotion-cognition interactions. J. Cogn. Neurosci. *18*, 871–879.

Beer, J.S., Knight, R.T., and D'Esposito, M. (2006b). Controlling the integration of emotion and cognition: the role of frontal cortex in distinguishing helpful from hurtful emotional information. Psychol. Sci. *17*, 448–453.

Bernoulli, D. (1954). Exposition of a new theory on the measurement of risk. Econometrica *22*, 23–36.

Boettiger, C.A., Mitchell, J.M., Tavares, V.C., Robertson, M., Joslyn, G., D'Esposito, M., and Fields, H.L. (2007). Immediate reward bias in humans: fronto-parietal networks and a role for the catechol-O-methyltransferase 158 (Val/Val) genotype. J. Neurosci. *27*, 14383–14391.

Chib, V.S., Rangel, A., Shimojo, S., and O'Doherty, J.P. (2009). Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. J. Neurosci. *29*, 12315–12320.

Ciaramelli, E., Muccioli, M., Làdavas, E., and di Pellegrino, G. (2007). Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. Soc. Cogn. Affect. Neurosci. *2*, 84–92.

Craig, A.D. (2002). How do you feel? Interoception: the sense of the physiological condition of the body. Nat. Rev. Neurosci. *3*, 655–666.

Cunningham, W.A., Johnson, M.K., Raye, C.L., Chris Gatenby, J., Gore, J.C., and Banaji, M.R. (2004). Separable neural components in the processing of black and white faces. Psychol. Sci. *15*, 806–813.

De Martino, B., Kumaran, D., Seymour, B., and Dolan, R.J. (2006). Frames, biases, and rational decision-making in the human brain. Science *313*, 684–687.

De Martino, B., Kumaran, D., Holt, B., and Dolan, R.J. (2009). The neurobiology of reference-dependent value computation. J. Neurosci. *29*, 3833–3842.

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., and Tsivkin, S. (1999). Sources of mathematical thinking: behavioral and brain-imaging evidence. Science *284*, 970–974.

Foot, P. (1978). The problem of abortion and the doctrine of double effect. In Virtues and Vices (Oxford: Blackwell).

Glimcher, P.W. (2009). Neuroeconomics and the study of valuation. In The Cognitive Neurosciences, Fourth Edition, M.S. Gazzaniga, ed. (Cambridge, MA: The MIT Press), pp. 1085–1092.

Greene, J.D. (2009). The cognitive neuroscience of moral judgment. In The Cognitive Neurosciences, Fourth Edition, M.S. Gazzaniga, ed. (Cambridge, MA: MIT Press), pp. 987–999.

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

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

Greene, J.D., Nystrom, L.E., Engell, A.D., Darley, J.M., and Cohen, J.D. (2004). The neural bases of cognitive conflict and control in moral judgment. Neuron *44*, 389–400.

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

Hare, T.A., O'Doherty, J.P., Camerer, C.F., Schultz, W., and Rangel, A. (2008). Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. J. Neurosci. *28*, 5623–5630.

Hare, T.A., Camerer, C.F., Knoepfle, D.T., and Rangel, A. (2010). Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. J. Neurosci. *30*, 583–590.

Hauser, M.D. (2006). The liver and the moral organ. Soc. Cogn. Affect. Neurosci. *1*, 214–220.

Hooker, C.I., and Knight, R. (2006). The role of lateral orbitofrontal cortex in the inhibitory control of emotion. In The Orbitofrontal Cortex, D.H. Zald and S.L. Rauch, eds. (New York: Oxford University Press), pp. 307–324.

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

Huebner, B., Dwyer, S., and Hauser, M. (2009). The role of emotion in moral psychology. Trends Cogn. Sci. *13*, 1–6.

Johnson, E.J., and Goldstein, D. (2003). Medicine. Do defaults save lives? Science *302*, 1338–1339.

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

Kim, S.H., and Hamann, S. (2007). Neural correlates of positive and negative emotion regulation. J. Cogn. Neurosci. *19*, 776–798.

Knutson, B., Taylor, J., Kaufman, M., Peterson, R., and Glover, G. (2005). Distributed neural representation of expected value. J. Neurosci. *25*, 4806–4812.

Knutson, B., and Greer, S.M. (2008). Anticipatory affect: neural correlates and consequences for choice. Philos. Trans. R. Soc. Lond. B Biol. Sci. *363*, 3771–3786.

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

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

Kringelbach, M.L., and Rolls, E.T. (2004). The functional neuroanatomy of the human orbitofrontal cortex: evidence from neuroimaging and neuropsychology. Prog. Neurobiol. *72*, 341–372.

Kuhnen, C.M., and Knutson, B. (2005). The neural basis of financial risk taking. Neuron *47*, 763–770.

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

Mohr, P.N.C., Biele, G., and Heekeren, H.R. (2010). Neural processing of risk. J. Neurosci. *30*, 6613–6619.

Moll, J., Krueger, F., Zahn, R., Pardini, M., de Oliveira-Souza, R., and Grafman, J. (2006). Human fronto-mesolimbic networks guide decisions about charitable donation. Proc. Natl. Acad. Sci. USA *103*, 15623–15628.

Montague, P., and Berns, G. (2002). Neural economics and the biological substrates of valuation. Neuron *36*, 265–284.

Nieder, A., and Miller, E.K. (2003). Coding of cognitive magnitude: compressed scaling of numerical information in the primate prefrontal cortex. Neuron *37*, 149–157.

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

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., and Dolan, R.J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. Science *304*, 452–454.

Padoa-Schioppa, C. (2007). Orbitofrontal cortex and the computation of economic value. Ann. NY Acad. Sci. USA *1121*, 232–253.

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

Petrinovich, L., and O'Neill, P. (1996). Influence of wording and framing effects on moral intuitions. Ethol. Sociobiol. *17*, 145–171.

Phan, K.L., Fitzgerald, D.A., Nathan, P.J., Moore, G.J., Uhde, T.W., and Tancer, M.E. (2005). Neural substrates for voluntary suppression of negative affect: a functional magnetic resonance imaging study. Biol. Psychiatry *57*, 210–219.

Platt, M.L., and Huettel, S.A. (2008). Risky business: the neuroeconomics of decision making under uncertainty. Nat. Neurosci. *11*, 398–403.

Price, D.D. (2000). Psychological and neural mechanisms of the affective dimension of pain. Science *288*, 1769–1772.

Qin, J., and Han, S. (2009). Parsing neural mechanisms of social and physical risk identifications. Hum. Brain Mapp. *30*, 1338–1351.

Rangel, A., Camerer, C., and Montague, P.R. (2008). A framework for studying the neurobiology of value-based decision making. Nat. Rev. Neurosci. *9*, 545–556.

Rule, R.R., Shimamura, A.P., and Knight, R.T. (2002). Orbitofrontal cortex and dynamic filtering of emotional stimuli. Cogn. Affect. Behav. Neurosci. *2*, 264–270.

Sanfey, A.G., Loewenstein, G., McClure, S.M., and Cohen, J.D. (2006). Neuroeconomics: cross-currents in research on decision-making. Trends Cogn. Sci. *10*, 108–116.

Schaich Borg, J., Hynes, C., Van Horn, J., Grafton, S., and Sinnott-Armstrong, W. (2006). Consequences, action, and intention as factors in moral judgments: an FMRI investigation. J. Cogn. Neurosci. *18*, 803–817.

Slovic, P. (2007). If I look at the mass I will never act: Psychic numbing and genocide. Judgment and Decision Making *2*, 79–95.

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

Thurstone, L.L. (1954). The measurement of values. Psychol. Rev. *61*, 47–58.

Tom, S.M., Fox, C.R., Trepel, C., and Poldrack, R.A. (2007). The neural basis of loss aversion in decision-making under risk. Science *315*, 515–518.

Tversky, A., and Kahneman, D. (1981). The framing of decisions and the psychology of choice. Science *211*, 453–458.

Ursu, S., Clark, K., Stenger, V., and Carter, C. (2008). Distinguishing expected negative outcomes from preparatory control in the human orbitofrontal cortex. Brain Res. *1227*, 110–119.

Van Essen, D.C. (2005). A population-average, landmark- and surface-based (PALS) atlas of human cerebral cortex. Neuroimage *28*, 635–662.

Venkatraman, V., Payne, J.W., Bettman, J.R., Luce, M.F., and Huettel, S.A. (2009). Separate neural mechanisms underlie choices and strategic preferences in risky decision making. Neuron *62*, 593–602.

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

Wager, T.D., Davidson, M.L., Hughes, B.L., Lindquist, M.A., and Ochsner, K.N. (2008). Prefrontal-subcortical pathways mediating successful emotion regulation. Neuron *59*, 1037–1050.

Wallis, J.D. (2007). Orbitofrontal cortex and its contribution to decision-making. Annu. Rev. Neurosci. *30*, 31–56.

Watanabe, M., and Sakagami, M. (2007). Integration of cognitive and motivational context information in the primate prefrontal cortex. Cereb. Cortex *17* (*Suppl 1*), i101–i109.

Yacubian, J., Sommer, T., Schroeder, K., Gläscher, J., Kalisch, R., Leuenberger, B., Braus, D.F., and Büchel, C. (2007). Gene-gene interaction associated with neural reward sensitivity. Proc. Natl. Acad. Sci. USA *104*, 8125–8130.

Moral judgments recruit domain-general valuation mechanisms to integrate representations of probability and magnitude.

Supplementary Information

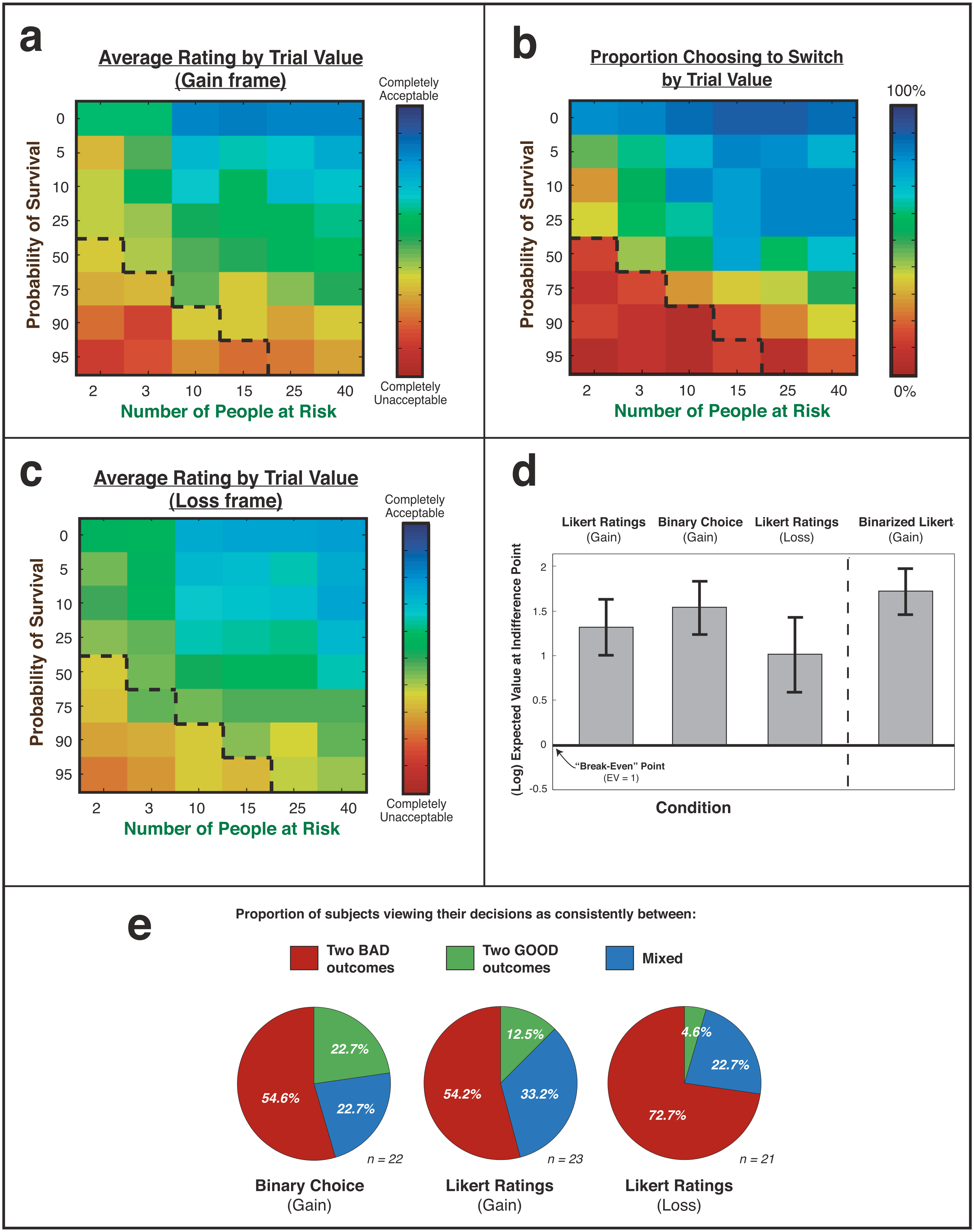
[1]Amitai Shenhav & [1]Joshua D. Greene

[1]Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, MA, 02138, USA

Correspondence should be addressed to A.S. (ashenhav@wjh.harvard.edu)

Supplementary Figures

Supplementary Figure 1 (related to Figure 2)**. a-c)** Average responses across trial value space for a replication of the fMRI experiment's behavioral task and two variations on that task, each with an independent group of subjects. Distribution and sampling of individual trial values was identical to the fMRI task in all versions (see Figure 2 and Methods). **(a)** Acceptability ratings (1-5 scale) replicating the behavioral results of the fMRI experiment, which employed a Gain frame, describing imperiled individuals as ones whose lives might be *saved*. **(b)** Behavioral results using binary choices instead of acceptability ratings (stay course = 0, change course = 1) and a Gain frame. **(c)** Behavioral results using acceptability ratings and a Loss frame, describing imperiled individuals as ones whose lives might be *lost*. d**)** Average log expected value for changing course at indifference point for responses shown in (a-c) as well as for a binarized version of (a). Average expected value is significantly higher than 1 (the expected value needed to break even; note that ln(1) = 0) in all cases (all p <0.05). Error bars indicate standard error of the mean. e) Distribution of subjects in each condition (a-c) classifying their options (one person vs group) as always consisting of two bad outcomes, two good outcomes, or mixed.


a
Average Rating by Trial Value
(Gain frame)
Completely Acceptable
Completely Unacceptable
Probability of Survival
0
5
10
25
50
75
90
95
2
3
10
15
25
40
Number of People at Risk
b
Proportion Choosing to Switch
by Trial Value
100%
0%
Probability of Survival
0
5
10
25
50
75
90
95
2
3
10
15
25
40
Number of People at Risk
c
Average Rating by Trial Value
(Loss frame)
Completely Acceptable
Completely Unacceptable
Probability of Survival
0
5
10
25
50
75
90
95
2
3
10
15
25
40
Number of People at Risk
d
Likert Ratings (Gain)
Binary Choice (Gain)
Likert Ratings (Loss)
Binarized Likert (Gain)
(Log) Expected Value at Indifference Point
2
1.5
1
0.5
0
-0.5
"Break-Even" Point (EV = 1)
Condition
e
Proportion of subjects viewing their decisions as consistently between:
🟥 Two BAD outcomes
🟩 Two GOOD outcomes
🟦 Mixed
22.7%
54.6%
22.7%
n = 22
Binary Choice (Gain)
12.5%
54.2%
33.2%
n = 23
Likert Ratings (Gain)
4.6%
22.7%
72.7%
n = 21
Likert Ratings (Loss)

Supplementary Tables

Supplementary Table 1 (related to Figure 6)**.** Regions exhibiting significant parametric increases in BOLD activity with linearly increasing tendency toward utilitarian judgment at the individual level.

Side	Region(s)	Cluster Extent	Peak Z score	Peak MNI Coordinates		
		(voxels)		X	Y	Z
B	SFG (medial wall)	347	4.66	-8	44	50
L	Lateral OFC / frontal pole	468	4.54	-46	36	-16
R	Anterior MTG, temporal pole	265	4.18	60	0	-32
R	Lateral OFC / frontal pole	411	4.14	42	38	-18
L	SPL, anterior IPS	273	3.86	-36	-36	44

Significant clusters met a voxelwise threshold of $p < .005$ and a cluster-wise threshold of $p < .05$. L, left; R, right; B, bilateral. SFG, superior frontal gyrus; OFC, orbitofrontal cortex; MTG, middle temporal gyrus; SPL, superior parietal lobule; IPS, intraparietal sulcus.

Supplementary Table 2 (related to Figure 7)**.** Regions exhibiting significant parametric increases in BOLD activity with increasing certainty of action outcome.

Side	Region(s)	Cluster Extent	Peak Z score	Peak MNI Coordinates		
		(voxels)		X	Y	Z
L	MTG, IPL	1839	5.89	-54	-48	0
L	vlPFC/IFG, aIns	2279	5.76	-50	16	16
L	MFG/premotor	383	4.62	-40	0	50
L	SFG (medial wall)	292	4.24	-8	42	44
B	SMA	304	4.20	-2	4	64
B	Precuneus, PCC	659	4.10	-8	-70	34
R	vlPFC, aIns	349	3.93	42	30	-4
R	IPL	268	3.78	44	-56	32

Significant clusters met a voxelwise threshold of $p < .005$ and a cluster-wise threshold of $p < .05$. L, left; R, right; B, bilateral. MTG, middle temporal gyrus; vlPFC, ventrolateral prefrontal cortex; IPL, inferior parietal lobule; IFG, inferior frontal gyrus; aIns, anterior insula; MFG, middle frontal gyrus; SFG, superior frontal gyrus; SMA, supplementary motor area; PCC, posterior cingulate cortex.

Supplemental Experimental Procedures

Task Instructions

Thank you for participating in this study. As noted on your consent form, any data collected from you will be kept strictly confidential.

In this study you will be asked to evaluate a number of moral dilemmas under various conditions. You will be presented with 10 different scenario contexts[1] and will respond to 10 conditions for each one.

For each scenario, you will proceed through four screens. The first three screens will start to describe a situation that you are hypothetically faced with and an action that you could perform in response to that situation.

[screen transition]

When you are done reading each screen, you should press any key to move on to the next one. However, please try your best to get the fullest understanding of the scenario as described thus far before moving on to the next screen.

The scenario description will include all the information you need to make your decision EXCEPT that it will not explicitly state the values for two features of the dilemma: a) a number of people involved in part of the scenario and b) a likelihood that something will happen. You will be evaluating this scenario given a number of variations of these features.

The fourth and final screen will provide you with the prompt that you will be answering for each of these variations - namely, whether or not it is morally acceptable for you to perform the action in question.[2]

[screen transition]

After you have understood the scenario context and the question you will be answering, you will press any button to move on to the individual trials that will fill these gaps for you in the scenario. Before you do so, please try your best to hold in mind what the action is that you will be evaluating across trials, as you will not be reminded of this after this screen.

First you will see a "+" in the middle of the screen. Any time that this is up, all you need to do is fixate on the "+" and prepare to respond to the next trial. Next you will see text appear indicating the missing information. You should then evaluate the action in question in the context of these values given, and make a judgment of its moral acceptability.

[1] In addition to the scenarios used for this experiment, subjects responded to a set of five additional scenarios (each as additional functional runs) for a separate experiment. These runs had a similar trial structure but the content of those scenarios and the trial value space from which the trials were drawn was different than those described in the current experiment. These additional runs were randomly ordered and interspersed with the one described here.

[2] In the binary choice follow-up experiment, this line read "The fourth and final screen will provide you with the prompt that you will be answering for each of these variations - namely, which of the actions in question you would perform in this scenario. It is important that you pay special attention to this as you will not be reminded later what specific actions you are choosing between."

You will rate each trial on a 1-5 scale, with 1 indicating that the action would be "Completely Unacceptable" and 5 indicating that it would be "Completely Acceptable."[3]

[screen transition]

After you answer you will again see a "+" in the middle of the screen, followed by the next trial. This will occur for 10 different variations of each scenario. It is important that you try your best to judge each trial in isolation, and avoid consideration of past responses for the current scenario or past scenarios.

You will only have 10 seconds to respond to each variation. If the "+" appears before you have responded, that means you are out of time. If this happens, do not attempt to respond. Simply look at the "+" and wait for the next trial.

Once you have pressed a button there is no way to go back to the previous screen. If you press the wrong button or if you press a button too soon, don't worry.

[screen transition]

Moral judgments can be difficult to make, and we understand that people sometimes change their minds about moral questions or feel conflicted about the answers they've given. Don't think of your answers as "written in stone." All we want from you is a thoughtful first response.

While we want your answers to be thoughtful, you may find that in some cases the right answer seems immediately obvious. If that happens, it's okay to answer quickly. There are no trick questions, and in every case we have done our best to make the relevant information as clear as possible.

Note, however, that no two scenarios are the same, although many are similar to each other. To answer a question properly you will have to read it carefully because it will always be different in some way from the questions you have already answered.

[screen transition]

In some cases, you might feel that the situation we've described is not realistic. For example, it might say that if you do X, then Y will happen, and you might think that this is not realistic, that Y might not necessarily happen if you do X. If you find yourself having these sorts of doubts, you should "suspend disbelief" and assume that the situation really is the way it's described, even if it doesn't seem realistic to you.

Likewise, you may feel that you need more information than is provided about the situation before you can give your answer. If this happens, you should make your best guess about what you think the situation is like without making any unnecessary assumptions. For example, if it doesn't say that the other person in the situation is related to you, then you should assume that you and the other person are unrelated.

[screen transition]

[3] In the binary choice experiment, these two lines read “You should then evaluate the actions in question in the context of these values given, and make your choice. You will press the 1 key to continue with your current action (STAY) or the 2 key to choose the alternate action (SWITCH).”

When you are done reading this screen you will begin two practice scenarios. Once again, you should use any key to advance through the first four screens. Then you will encounter a number of trials which you will rate on a 1-5 scale.

While you will see 10 such trials for each scenario in the actual experiment, you will only respond to 4 variations of each practice scenario.

Furthermore, when the actual experiment begins there will be an additional 10 seconds of fixation before the scenario text comes up, while the scanner warms up. Please begin reading as usual when the text comes up.

If you have any questions at this time, please wait until the current scan completes and then squeeze the ball to let the experimenter know. Otherwise, you may press a key to move onto the practice.

Dilemma text

Below is the text for the moral dilemmas used in the present study (including the wording used to describe the variable Magnitude and Probability parameters). See Fig 1 and Methods in the main text for details concerning presentation. Dilemmas have been adapted from a variety of sources(Boorse and Sorensen, 1988; Cushman et al., 2006; Foot, 1978; Greene et al., 2001; Royzman and Baron, 2002; Thomson, 1985; Waldmann and Dieterich, 2007).

Practice dilemmas

Skydivers

You are the head of a safety crew that has been alerted about a group of skydivers that is wearing defective parachutes and is currently in a freefall. You've arrived in time to see that a group of divers is holding onto one another and will fall to their death a few feet to your right unless your crew's trampoline is moved beneath them.

Your crew sees one diver apart from the rest who is much closer to reaching the ground, and instinctively moves a few feet to the left to catch him. You know that there won't be enough time to reach the group if you wait to catch the one. If nothing is done, the safety crew will catch the diver to the left, avoiding his death, but the separate group of divers will land on the ground, causing their deaths. You can re-direct your crew to the right, preventing the death of the group, but the one diver won't be caught.

You also know that there is a chance that one of the parachutes of the group will still open in time to prevent all of their deaths, but there is no chance that this will happen for the one given how close he is to the ground.

Consider each of the following scenarios and, for each one, determine how morally acceptable you think it would be to direct your crew to the right, re-positioning the trampoline underneath the group.

Divers on right: [variable Magnitude value appears here]

Probability of parachute opening: [variable Probability value appears here]

Loss Frame: *Probability of continuing freefall: [variable 100-Probability value appears here]*

Scuba Divers (only included in fMRI task)

You are working on a ship when you see a large group of sharks in the distance heading towards a team of scuba divers. You know that when the sharks reach this group of people they will encircle them and kill them all.

You also know that there is one scuba diver separate from this group who is closer to you and currently not in the path of the sharks. If you shoot this one diver with a harpoon, he will certainly die and the sharks will sense his presence and switch trajectory away from the team of divers. If you do this, all of the divers will be spared.

You also know that the divers are carrying an electromagnetic (EM) pulse that will deter the sharks, but given the team's current depth there is only a small chance that the pulse will function. In order to ensure that the group will be spared, you must shoot the one diver now.

Consider each of the following scenarios and, for each one, determine how morally acceptable it would be for you to shoot the one diver.

Scuba divers in team: [variable Magnitude value appears here]

Probability EM pulse will function: [variable Probability value appears here]

Main experiment dilemmas

Gas Vent

There has been an accident in the basement of an office building. A pipe has broken and is releasing a lethal gas into the building's ventilation system. If nothing is done, the gas will travel upward to Room A and cause the death of all workers there.

You are the building manager. You know that the only way to avoid the deaths of the workers in Room A is to block the main vent immediately by hitting a button that will close one of the vent doors. You also know that if you close the vent door the gas will be diverted into a different room, Room B, and cause the death of the one worker there.

You are also aware that there is a chance that the workers in Room A will escape before the gas reaches them. This would be impossible for the one in Room B.

Consider each of the following scenarios and, for each one, determine how morally acceptable you think it would be to hit the button that closes the vent door and diverts the lethal gas to Room B.

Workers in Room A: [variable Magnitude value appears here]

Probability of escape: [variable Probability value appears here]

Loss Frame: *Probability of remaining trapped: [variable 100-Probability value appears here]*

Bridge

You are operating a crane in front of a tall bridge when an earthquake hits. As a result of the earthquake, a gap has formed in the bridge, and a bus carrying a number of people is now partially hanging off of this gap in the bridge. If nothing is done, the bus will soon plummet and all of these people will be killed.

You know that you can move your crane in a way to push the bus back onto the bridge, avoiding the deaths of all the people on board. However, you are also aware that in its current position your crane is supporting one small part of the bridge on which there is currently a single car. If you move the crane, this part will certainly collapse and the one person in that car will be killed.

You also know that there is a chance the bus will be able to drive back onto the bridge, preventing the death of everyone on board. There is no chance that the car will be able to avoid falling if you move your crane.

Consider each of the following scenarios and, for each one, determine how morally acceptable you think it would be for you to move your crane to push the bus, causing part of the bridge with a single car to collapse.

People on bus: [variable Magnitude value appears here]

Probability bus will drive to safety: [variable Probability value appears here]

Loss Frame: *Probability bus will fall off bridge: [variable 100-Probability value appears here]*

Rescue Boat

You are driving a rescue boat in the ocean, heading east towards one drowning man. You receive a distress signal informing you that a small boat has capsized in the opposite direction, and all the people aboard are now drowning.

You know that if you immediately change course and go full speed, bearing west, you will reach these people in time to save them. However, if you do this, the one man to the east will certainly die. If you do nothing and hold your course, the one man will be saved, but you will not reach the people to the west in time to save them.

You also know that the only other rescue boat in the area is much further to the west, so would be unable to reach the one drowning man. But there is a chance the rescue boat will reach the group drowning to the west.

Consider each of the following scenarios and, for each one, determine how morally acceptable you think it would be to change your course to head toward the group to the west.

Individuals drowning to the west: [variable Magnitude value appears here]

Probability of alternate rescue boat reaching them: [variable Probability value appears here]

Loss Frame: *Probability of alternate rescue boat failing to reach them: [variable 100-Probability value appears here]*

Boxcar

You are operating the switch at a railroad station when you see an empty, out of control boxcar coming down the main track. It is moving so fast that anyone it hits will die immediately. The boxcar is headed towards a tunneled section in which a group of repairmen are working.

You can flip the switch, redirecting the boxcar to a sidetrack on which there is one repairman working. If you do nothing, the boxcar will continue toward the repairmen in the tunnel on the main track and kill them all. If you hit the switch, the repairmen on the main track will be spared but the one repairman on the sidetrack will be hit by the boxcar and die.

You know that there is a chance an alarm on the main track will be triggered in time to alert the repairmen to evacuate before the boxcar arrives. There is no such alarm on the sidetrack, and therefore no chance the one workman would evacuate in time.

Consider each of the following scenarios and, for each one, determine how morally acceptable you think it would be to hit the switch, redirecting the boxcar onto the sidetrack.

Repairmen on main track: [variable Magnitude value appears here]

Probability of evacuation: [variable Probability value appears here]

Loss Frame: *Probability of remaining in tunnel: [variable 100-Probability value appears here]*

Cafe Grenade

You are working in the kitchen of a café, and see a terrorist throw a grenade next to the main dining room, in which a number of customers are eating. If nothing is done the grenade will explode and the walls of the main dining room will collapse and kill these customers.

There is only one other location the grenade could be thrown before it explodes, and that is the patio outside of the main dining room. There is one customer sitting on the patio, and he would certainly be killed if the grenade is thrown there.

You have time to run out and throw the grenade to the patio, without risking any harm to yourself. If you do this, the customers in the dining room will live but the one on the patio will die. You also know that there is a chance that the dining room walls will withstand the blast of the grenade explosion, and the customers would be spared.

Consider each of the following scenarios and, for each one, determine how morally acceptable you think it would be to throw the grenade onto the patio.

Customers in main dining room: [variable Magnitude value appears here]

Probability customers will be spared: [variable Probability value appears here]

Loss Frame: *Probability walls will collapse on customers: [variable 100-Probability value appears here]*

Post-Task Survey (only used in behavioral experiments subsequent to fMRI task; only final question was included in analysis, wherein the scale was split into 3 even parts such that: 1-3 was categorized as “consistently two bad outcomes”; 4-6 as “mixed”; and 7-9 as “consistently two good outcomes”))

Please circle a number to indicate how much you agree with each of the following statements about what motivated your earlier judgments:

1. My judgments were driven overall by a positive feeling towards saving the one person.

Strongly Disagree	1	2	3	4	5	6	7	8	9	**Strongly Agree**

2. My judgments were driven overall by a positive feeling towards saving the group.

Strongly Disagree	1	2	3	4	5	6	7	8	9	**Strongly Agree**

3. My judgments were driven overall by a negative feeling towards letting the one person die.

Strongly Disagree	1	2	3	4	5	6	7	8	9	**Strongly Agree**

4. My judgments were driven overall by a negative feeling towards letting the members of the group die.

Strongly Disagree	1	2	3	4	5	6	7	8	9	**Strongly Agree**

When judging the scenarios you read earlier, people can have two different kinds of reactions to their options. Some people feel as though the choice they are endorsing is primarily one of **two bad outcomes** (the “lesser of two evils”) because of the deaths that may occur either way. Others feel like they are primarily choosing between **two good outcomes** because either choice offers the opportunity to save lives.
Which of these better describes how you felt about your choices?

Always felt like two BAD outcomes	1	2	3	4	5	6	7	8	9	**Always felt like two GOOD outcomes**

Supplemental References

Boorse, C., and Sorensen, R. (1988). Ducking harm. Journal of Philosophy *85*, 115-134.

Cushman, F., Young, L., and Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: testing three principles of harm. Psychol Sci *17*, 1082-1089.

Foot, P. (1978). The problem of abortion and the doctrine of double effect. In Virtues and Vices (Oxford: Blackwell).

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

Royzman, E.B., and Baron, J. (2002). The preference for indirect harm. Social Justice Research *15*, 165-184.

Thomson, J. (1985). The Trolley Problem. Yale Law Journal *94*, 1395-1415.

Waldmann, M.R., and Dieterich, J.H. (2007). Throwing a bomb on a person versus throwing a person on a bomb: intervention myopia in moral intuitions. Psychol Sci *18*, 247-253.