You are driving a rescue boat in the ocean, heading east towards the people aboard are now drowning. That a small boat has capsized in the opposite direction, and all five of those people are now drowning to the west. You also know that the only other rescue boat in the area is much faster, speeding, bearing west, you will reach these people in time to save them. However, if you do this, the one man to the east will be saved, but you will not reach the people to the west in time to save them. You will simply watch the man to the east drown — he will be left to drown in the water. The people to the west will otherwise die. Determining the acceptability of an action in such cases must then rely to some degree on an evaluation of expected gains and/or losses of lives that would result.

Behavioral Results
When entered into a mixed-effects multiple regression, subjects’ ratings of moral acceptability were shown to be highly sensitive to Mag (t(34,1) = 9.71), Prob (t(34,1) = 15.3), and EV (t(34,1) = 4.29) all p-values < 0.0001)

This is especially noteworthy given that these values were randomly distributed between very different dilemmatic contexts.

Subjects’ trial RT’s, however, were shown not to be significantly influenced by Mag and Prob (t(34,1) = 1.57, p=0.13, t(34,1) = 1.26, p=0.22), but were faster as increased EV (t(34,1) = 3.98, p<0.0005)

The scenario descriptions left unstated the probability of survival.

Ratings by Trial Value (Group Average) (natural-log transformed based on best fit to normal distribution)

Behavioral Methods
35 Subjects (16 male, ages 18-42) each performed five fMRI runs. Whole-brain fMRI was performed on a 3T Siemens Trio magnet with the following parameters:
- TR = 2.5s, TE = 28ms, FOV = 256x256mm, Matrix = 96x96
- 42 slices, thickness = 3mm, gap = 0.5mm
- Data were concatenated across runs within each subject and trials were analyzed using a variable-duration canonical (SPM) HRF within a multiple regression model that included regressors for reaction time (to correct for differences in time-on-task) as well as:
  a) Mag (natural-log transformed based on best fit to behavioral data)
  b) Prob (coded as probability of group NDT surviving if prescribed action isn’t taken, shown as positive direction in x-axis on left)
  c) EV (multiplicative interaction of Mag and Prob as shown in color map on left)

Neuroimaging Results
Below are the statistical maps (thresholded for cluster size of 20 voxels) for positively increasing BOLD activity for each of the 3 regressors of interest:

Magnitude (numbers of people who could be saved/killed)
- Bilateral dorsal striatum (BdStr) (caudate nucleus and putamen) and central insula (cIns)
- Bilateral middle frontal gyri (MFG)
- Anterior and posterior cingulate cortices (ACC, PCC/premotor)

Probability (the probability that the group will die if action is not taken)
- Left dorsal posterior insula (dPI)
- Right anterior/central insula (a/PI)
- B-PI

Expected Value (overall value of allowing one person to die to save the group)
- Ventromedial prefrontal cortex (vmPFC)
- Left ventrolateral prefrontal cortex (vLPI)
- Right posterior orbitofrontal cortex (rPFC)

Expected Value

Conclusions
1) Subjects were highly sensitive to “expected moral value” and its components (the number of lives saved and probability of death) when making judgments of the moral acceptability of sacrificing the life of another.
2) Brain regions whose activations were differentially correlated with increases in each of these parameters shared a number of commonalities with those implicated in decisions involving real reward or loss for the subject. This is despite the fact that our task involved judgments that were a) hypothetical and b) concerned only with effects on individuals other than the subject.

Behavioral Task
Subjects were presented with 5 different scenarios in which they evaluated the moral acceptability of a personal gain or loss.

Background
Neuroeconomic studies have sought to isolate brain regions responsible for representing risk, reward magnitude, and overall expected value in order to inform affective judgments. Almost all of these studies have utilized stimuli that involve the opportunity for (tangible) personal gain or loss.

Trolley-type moral dilemmas present individuals with choices in which promoting the “greater good” (maximizing the number of lives saved) requires causing the death of another (e.g., pulling a switch that causes a trolley to run over one worker rather than the five it would otherwise hit). Determining the acceptability of an action in such cases must then rely to some degree on an evaluation of expected gains and/or losses of lives that would result.

Purpose
To determine which brain regions are involved in representing magnitude, probability, and expected value in the context of moral dilemmas in which the decision-maker is in an unaffected third party with no material stake in the outcome.

Neuroimaging Methods
35 Subjects (16 male, ages 18-42) each performed five fMRI runs. Whole-brain fMRI was performed on a 3T Siemens Trio magnet with the following parameters:
- TR = 2.5s, TE = 28ms, FOV = 256x256mm, Matrix = 96x96
- 42 slices, thickness = 3mm, gap = 0.5mm
- Data were concatenated across runs within each subject and trials were analyzed using a variable-duration canonical (SPM) HRF within a multiple regression model that included regressors for reaction time (to correct for differences in time-on-task) as well as:
  a) Mag (natural-log transformed based on best fit to behavioral data)
  b) Prob (coded as probability of group NDT surviving if prescribed action isn’t taken, shown as positive direction in x-axis on left)
  c) EV (multiplicative interaction of Mag and Prob as shown in color map on left)

Neuroimaging Results
Below are the statistical maps (thresholded for cluster size of 20 voxels) for positively increasing BOLD activity for each of the 3 regressors of interest:

Magnitude (numbers of people who could be saved/killed)
- Bilateral dorsal striatum (BdStr) (caudate nucleus and putamen) and central insula (cIns)
- Bilateral middle frontal gyri (MFG)
- Anterior and posterior cingulate cortices (ACC, PCC/premotor)

Probability (the probability that the group will die if action is not taken)
- Left dorsal posterior insula (dPI)
- Right anterior/central insula (a/PI)
- B-PI

Expected Value (overall value of allowing one person to die to save the group)
- Ventromedial prefrontal cortex (vmPFC)
- Left ventrolateral prefrontal cortex (vLPI)
- Right posterior orbitofrontal cortex (rPFC)

Relevance to Decision-Making Literature
- Studies in both humans and non-human primates have shown the mPFC and OFC to play a role in risk-informed expected value of both primary and secondary reinforcers and integrating the determinative value in present and future behavior.
- Knutson et al., 2005; Tom et al., 2007

Conclusions
1) Subjects were highly sensitive to “expected moral value” and its components (the number of lives saved and probability of death) when making judgments of the moral acceptability of sacrificing the life of another.
2) Brain regions whose activations were differentially correlated with increases in each of these parameters shared a number of commonalities with those implicated in decisions involving real reward or loss for the subject. This is despite the fact that our task involved judgments that were a) hypothetical and b) concerned only with effects on individuals other than the subject.

References
Knutson et al., 2005; Tom et al., 2007

Relevance to Decision-Making Literature
- Studies in both humans and non-human primates have shown the mPFC and OFC to play a role in risk-informed expected value of both primary and secondary reinforcers and integrating the determinative value in present and future behavior.
- Knutson et al., 2005; Tom et al., 2007

Conclusions
1) Subjects were highly sensitive to “expected moral value” and its components (the number of lives saved and probability of death) when making judgments of the moral acceptability of sacrificing the life of another.
2) Brain regions whose activations were differentially correlated with increases in each of these parameters shared a number of commonalities with those implicated in decisions involving real reward or loss for the subject. This is despite the fact that our task involved judgments that were a) hypothetical and b) concerned only with effects on individuals other than the subject.