Supporting Online Material for

The Neural Basis of Loss Aversion in Decision-Making Under Risk

Sabrina M. Tom, Craig R. Fox, Christopher Trepel, Russell A. Poldrack

*To whom correspondence should be addressed. E-mail: poldrack@ucla.edu

Published 26 January 2007, *Science* **315**, 515 (2007)

DOI: 10.1126/science.1134239

This PDF file includes:

Supplementary materials for:

Losses loom larger than gains in the brain: Neural loss aversion predicts behavioral loss aversion

Sabrina M. Tom[1], Craig R. Fox[1,2], Christopher Trepel[2], & Russell A. Poldrack[1,3,4]

1. Department of Psychology, UCLA
2. Anderson School of Management, UCLA
3. Brain Research Institute, UCLA
4. Department of Psychiatry and Biobehavioral Sciences, UCLA

Materials and Methods

Participants: Sixteen right-handed, healthy, English-speaking participants (nine females; mean age, 22 ± 2.9 years) were recruited through advertisements posted on the UCLA campus. All participants were free of neurological and psychiatric history and gave informed consent to participate according to a protocol approved by the University of California, Los Angeles Institutional Review Board.

Pre-testing and endowment session. Prior to fMRI scanning, participants were endowed with a $30 cash payment for their participation in an initial pre-testing session. The payment was made at least a week in advance of the scanning session in order to minimize the potential risk-seeking that can occur in response to windfall gains (i.e., when one is "playing with the house money") (cf. S*1*). During this session, they were presented with a questionnaire regarding gambling attitudes and made a number of choices involving hypothetical gambles.

Scanning session. In order to convince participants that this was a real gambling experiment in which they would be exposed to a real possibility of losing their own money, we asked them to bring $60 in cash with them on the day of the scan and told them that this was the maximum amount that they could possibly lose. In actuality, due to the positive expected value of the gambles that participants evaluated, such a negative outcome was highly unlikely, and in fact no participant lost more than $12 from these gambles. The average amount won was $23. Ten participants won money (max gain= $70) and three participants lost money (max loss= $12) from gambling. The remaining three participants rejected all three trials that were selected, and thus received no additional money. Due to the initial $30 endowment, all participants left the experiment with a net gain, ranging from $18 – $100.

In the scanner, participants were presented with 3 runs of 85-86 trials, each of which proposed a mixed gamble entailing a 50/50 chance of gaining one amount of money or losing another amount. Possible gains ranged from $10-$40 (in $2 increments) and possible losses ranged from $5-$20 (in $1 increments). All 256 possible combinations of

gains and losses were presented across the three runs. Participants were asked to evaluate whether or not they would like to play each of the gambles presented to them. They were told that one trial from each of the runs would be selected at random, and if they had accepted that gamble during scanning, the outcome would be decided with a coin toss; if they had rejected the gamble, then the gamble would not be played.

In order to encourage participants to reflect on the subjective attractiveness of each gamble rather than revert to a fixed decision rule (e.g., accept only if gain $\geq 2 \times$ loss), we asked them to indicate one of four responses to each gamble (strongly accept, weakly accept, weakly reject, and strongly reject) using a four-button response box. We also instructed them to respond as quickly as possible within the 3-second trial duration.

Stimulus presentation and timing of all stimuli and response events were achieved using Matlab and the Psychtoolbox (www.psychtoolbox.org) on an Apple PowerBook running Mac OS 9 (Apple Computers, Cupertino, CA). Visual stimuli were presented using MRI-compatible goggles (Resonance Technologies, Van Nuys, CA). The timing and order of stimulus presentatation was optimized for estimation efficiency using optseq2 (http://surfer.nmr.mgh.harvard.edu/optseq/) (S*2*).

Behavioral analysis. Statistical analyses of behavioral data were performed using the R statistical package (http://www.r-project.org). Logistic regression was performed on the behavioral data after collapsing strong/weak responses into accept and reject categories, with the size of the potential gain and loss as independent variables and acceptance/rejection as the dependent variable. This analysis was performed separately for each participant, collapsing over scanning runs. Behavioral loss aversion (λ) was computed as:

$$\lambda = -\beta_{loss} / \beta_{gain}$$

where β_{loss} and β_{gain} are the unstandardized regression coefficients for the loss and gain variables, respectively. This parameter is similar to the λ parameter in prospect theory (S*3*) but makes the common simplifying assumptions of a linear rather than curvilinear value function, and identical decision weights for a 0.5 probability to gain or lose money.

MRI data acquisition. Imaging was performed using a 3T Siemens AG (Erlangen, Germany) Allegra MRI scanner at the UCLA Ahmanson-Lovelace Brain Mapping Center. We acquired 240 functional T2*-weighted echoplanar images (EPI) [slice thickness, 4 mm; 34 slices; repetition time (TR), 2 s; echo time (TE), 30 ms; flip angle, 90°; matrix, 64 x 64; field of view (FOV), 200 mm]. Two additional volumes were discarded at the beginning of each run to allow for T1 equilibrium effects. In addition, a T2-weighted matched-bandwidth high-resolution anatomical scan (same slice prescription as EPI) and magnetization-prepared rapid-acquisition gradient echo (MPRAGE) were acquired for each subject for registration purposes (TR, 2.3; TE, 2.1; FOV, 256; matrix, 192 x 192; sagittal plane; slice thickness, 1 mm; 160 slices). The orientation for matched-bandwidth and EPI scans was oblique axial so as to maximize full brain coverage and to

optimize signal from ventromedial prefrontal regions.

Imaging preprocessing and registration. Initial analysis was performed using the FSL toolbox from the Oxford Centre for fMRI of the Brain (www.fmrib.ox.ac.uk/fsl). The image timecourse was first realigned to compensate for small head movements (S*4*). Translational movement parameters never exceeded 1 voxel (3.125 mm inplane, 4 mm throughplane) in any direction for any subject or session. In cases where translational motion of more than 1 mm was detected in any direction, images were denoised using MELODIC independent components analysis within FSL (this was performed for 22 runs in 9 participants). Motion-related components were identified manually using a set of heuristics (Poldrack, Aron, & Tom, 2005, Human Brain Mapping Abstracts), and the data were then reconstituted after removing the motion-related components. Data were spatially smoothed using a 5 mm full-width-half-maximum Gaussian kernel. Registration was conducted through a 3-step procedure, whereby EPI images were first registered to the matched-bandwidth high-resolution structural image, then to the MPRAGE structural image, and finally into standard [Montreal Neurological Institute (MNI)] space (MNI avg152 template), using 12-parameter affine transformations (S*5*). Statistical analyses were performed in native space, with the statistical maps normalized to standard space prior to higher-level analysis.

Statistical analysis. Whole-brain statistical analysis was performed using a multi-stage approach to implement a mixed-effects model treating participants as a random effect. Statistical modeling was first performed separately for each imaging run. Regressors of interest were created by convolving a delta function representing trial onset times with a canonical (double-gamma) hemodynamic response function.

For the primary whole-brain analyses, two modeling approaches were used. In the first (referred to as the "parametric analysis"), all trials were modeled using a single condition (i.e., overall task-related activation; see Figure S1), and three additional orthogonal parametric regressors were included representing: (a) the size of the potential gain (see Figure S2), (b) the size of the potential loss (see Figure S3), and (c) the Euclidean distance of the gain/loss combination from the diagonal of the gamble matrix (i.e., distance from indifference assuming lambda=2 and a linear value function). This latter variable was included because of behavioral evidence suggesting greater difficulty making a decision for trials in which participants had the weakest preference (See Figure 1C in main text), however, these results are not discussed in the present paper; a fuller account of this latter analysis will be presented elsewhere. In the second approach (referred to as the "matrix analysis"), the gain/loss matrix was collapsed from 16 × 16 into a 4 × 4 matrix (see Figure 1 in the main text), and trials from each of the 16 resulting cells were modeled as separate conditions. This allowed separate estimation of the evoked response for each of these cells at each voxel; the primary use of the matrix analysis was to create the heatmaps of activation presented in Figure 2, and to perform the comparison between best versus worst gambles described in the main text. For all analyses, time-series statistical analysis was carried out using FILM (FMRIB's Improved Linear Model) with local autocorrelation correction (S*6*) after highpass temporal filtering (Gaussian-weighted LSF straight line fitting, with sigma=33.0s).

For each of these lower-level analyses, a higher-level analysis was performed that combined all sessions for each participant using the FMRIB Local Analysis of Mixed Effects (FLAME) module in FSL (S*7,* S*8*), and a one-sample t-test was performed at each voxel for each contrast of interest. Z (Gaussianised T) statistic images were thresholded using clusters determined by $Z > 2.3$ and a (whole-brain corrected) cluster significance threshold of $p < .05$ using the theory of Gaussian Random Fields (S*9*). For the comparison between best versus worst gamble conditions, control for multiple comparisions was implemented using randomization tests (S*10*) with the FSL randomize tool in order to allow correction limited to small regions of interest.

For whole-brain analyses of correlations between neural activity and behavioral parameters across participants (Figures S4-7), voxelwise robust regression was used to in order to reduce the influence of outliers on the analysis (cf. S*11*). Because Gaussian random field results were not available for these analyses, whole-brain correction for multiple comparisons was implemented by controlling the false discovery rate (FDR) at $q < .05$ (S*12*) along with a cluster extent threshold of 100 voxels. Because FDR is adaptive, the t-threshold that controls FDR at $q < 0.05$ varies between analyses.

Renderings. All statistical maps are presented at a whole-brain corrected significance level of $p < .05$, either using GRFT or FDR corrections, and are overlaid on a group mean structural image. Cortical renderings were performed using CARET software (http://brainmap.wustl.edu). Group statistical maps were mapped into to the Probabilistic Average Landmark and Surface-based (PALS) atlas using the multifiducial mapping technique described by Van Essen (S*13*). For the purposes of presentation, data are overlaid on the average atlas surface.

Conjunction analysis. Conjunction analysis for gains and losses (Figure 2 and Table S1) was performed by multiplying binarized versions of the thresholded statistical maps obtained for the parametric gain and loss analyses. Because each of these maps is itself whole-brain corrected at $p < .05$, this conjunction tests against the conjunction null at $p < .05$ (S*14*).

Computation of neural loss aversion. Because the neural gain and loss coefficients were broadly distributed and spanned zero, it was not possible to compute a stable loss aversion coefficient in the same way used for the behavioral data (i.e., the ratio of loss to gain responses). Instead, we computed neural loss aversion at every voxel by subtracting the slope of the gain response from the (negative) slope of the loss response. Whole-brain analyses using the resulting images were performed using robust regression with false discovery rate correction (Figure S6 and Table S2).

Region-of-interest (ROI) analyses. For the purposes of exploratory analysis, ROIs were created based on the significant clusters of activation in the voxelwise analyses. Using these regions of interest, ROI analyses were performed by extracting parameter estimates (betas) from the fitted model and averaging across all voxels in the cluster for each subject. For analyses of correlations between behavioral and ROI data, robust regression was used to minimize the impact of outliers in the behavioral data, using iteratively re-

weighted least squares implemented in the robustfit command in the MATLAB Statistics Toolbox. Reported r-values reflect (non-robust) Pearson product-moment correlation values, whereas the reported p-values and regression lines are based on the robust regression results.

Mediation analysis. Simple mediation analysis was performed on data from each of the significant clusters in the neural loss aversion analysis (listed in Table S2) as described by Preacher and Hayes (S*15*). Because of the small sample size, 95% bias-corrected and accelerated confidence intervals were generated for the indirect effect using bootstrapping with the R software package. The foregoing analysis demonstrates a direct relation between neural and behavioral loss aversion. We further investigated whether this relationship was driven more by the neural processing of potential gains or losses in each of the clusters identified in the foregoing analysis. Whereas all of the regions had significant negative correlations between behavioral loss aversion and the (negative) neural loss response, only one region showed a significant relationship between behavioral loss aversion and the neural gain response (see Table S2). Thus, in agreement with the whole-brain analysis above, participants who were less behaviorally loss averse (i.e., more risk seeking) were less neurally sensitive to the size of the potential loss. In order to more directly characterize the relative roles of gain and loss responses in behavioral loss aversion, the data from each cluster were entered into a mediation analysis, which revealed that the effect of the neural gain response on (log) behavioral λ was mediated by the neural loss response in five of the eight clusters (including bilateral lateral PFC, pre-SMA, and ventral striatum).

Figure S1. Regions with significant activation for task vs. baseline ($Z > 2.3$, whole-brain cluster-corrected at $p < .05$ using GRFT). Red-yellow scale reflects positive activation, blue-white scale reflects negative activation.

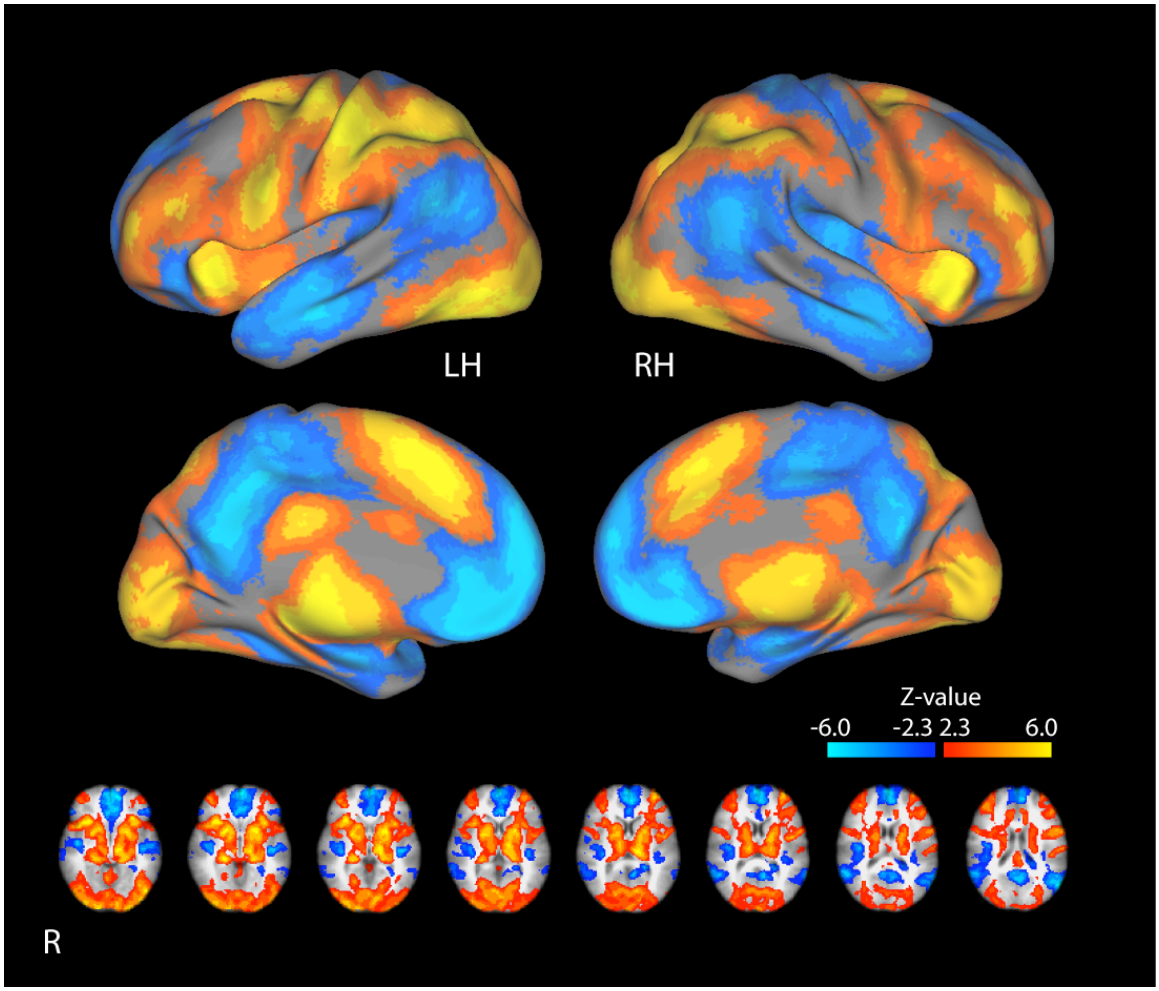

Figure S2. Regions with significant parametric increase in fMRI signal to increasing potential gains ($Z > 2.3$, whole-brain cluster-corrected at $p < .05$ using GRFT). No regions showed decreasing activity for increasing potential gains.

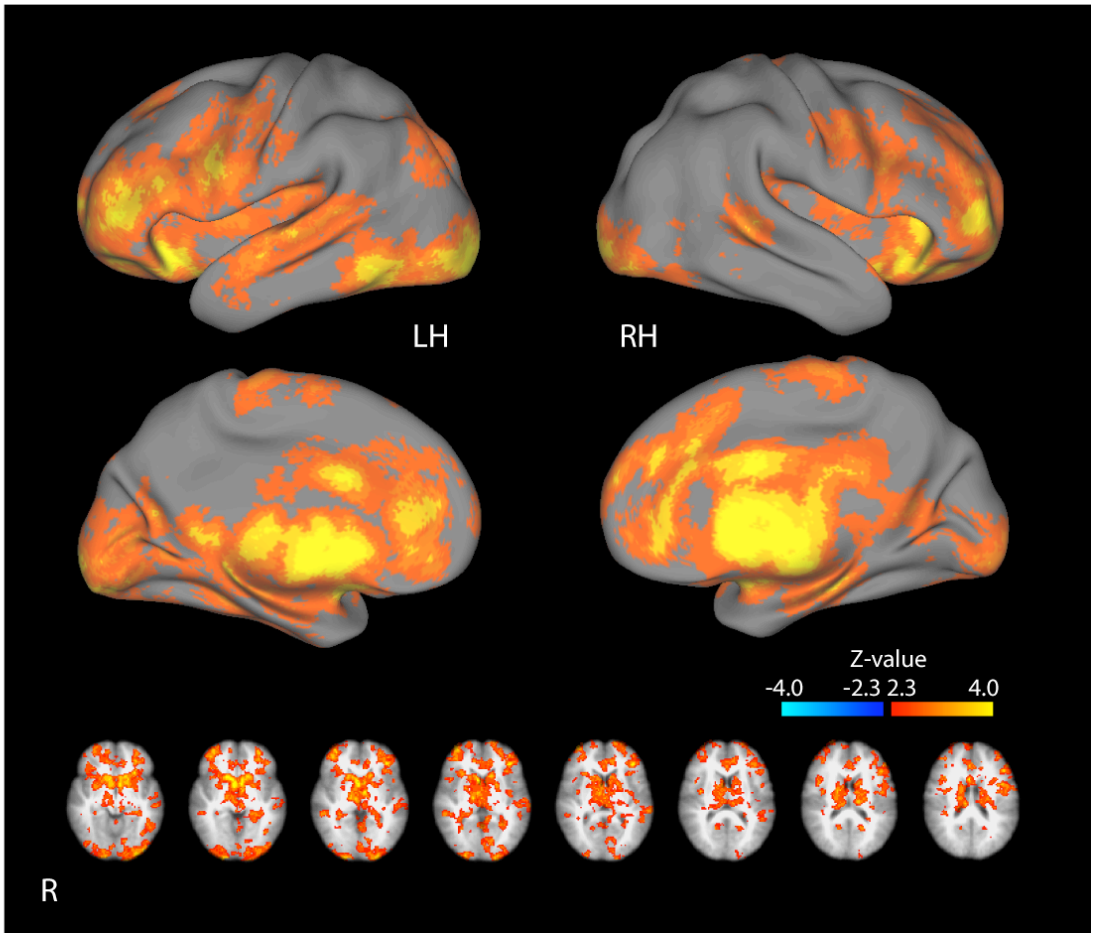

Figure S3. Regions with significant parametric decrease to increasing potential losses (Z > 2.3, whole-brain cluster-corrected at $p < .05$ using GRFT). No regions showed increasing activity for increasing potential losses.

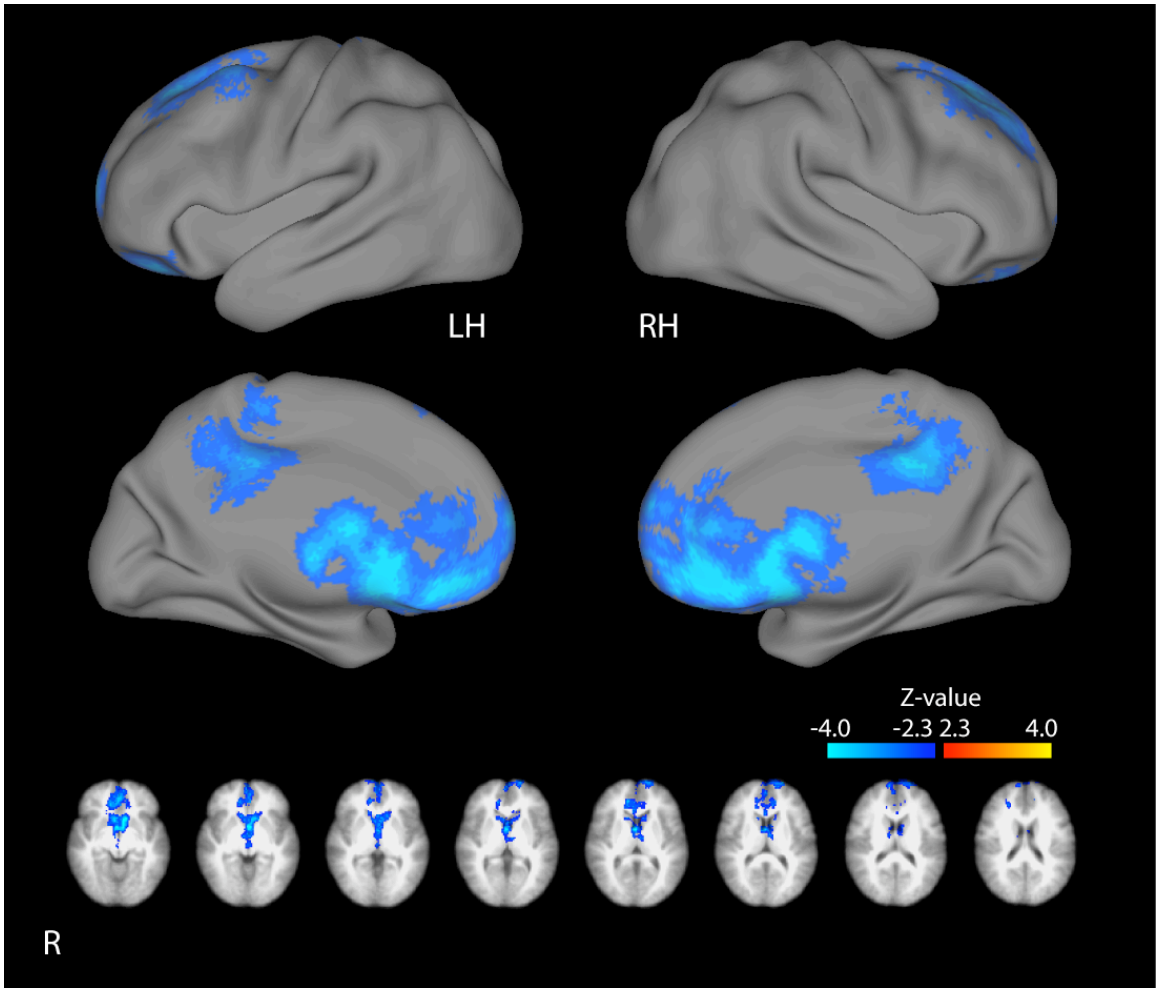

Figure S4. Regions with significant positive correlation between the parametric response to potential gains and behavioral loss aversion (ln(λ)) across participants (whole brain false discovery rate corrected at $q < 0.05$ [$t > 4.3$] and cluster extent > 100 voxels). No regions showed significant negative correlation.

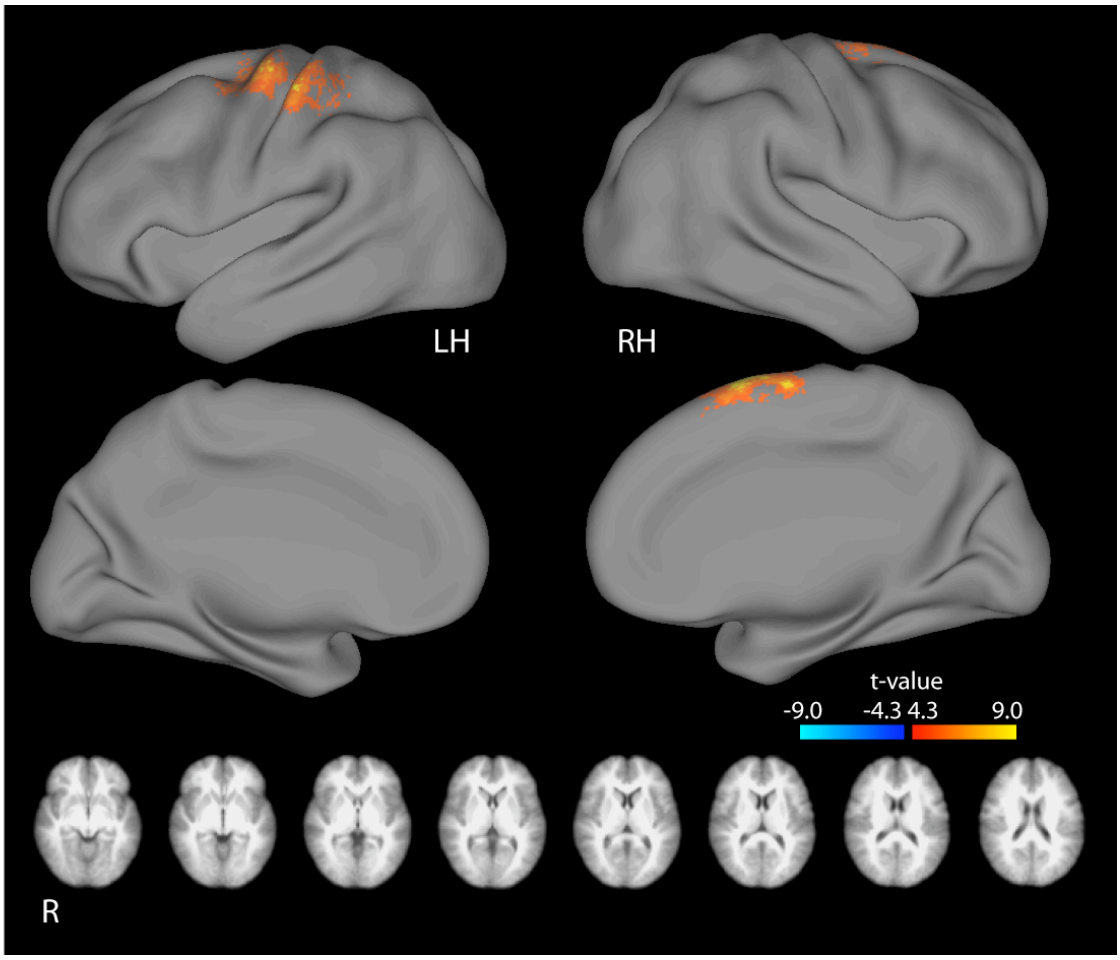

Figure S5. Regions with significant positive correlation between the parametric response to potential losses and behavioral loss aversion (ln(λ)) across participants (whole brain false discovery rate corrected at $q < 0.05$ [$t > 3.1$] and cluster extent > 100 voxels). No regions showed significant positive correlation.

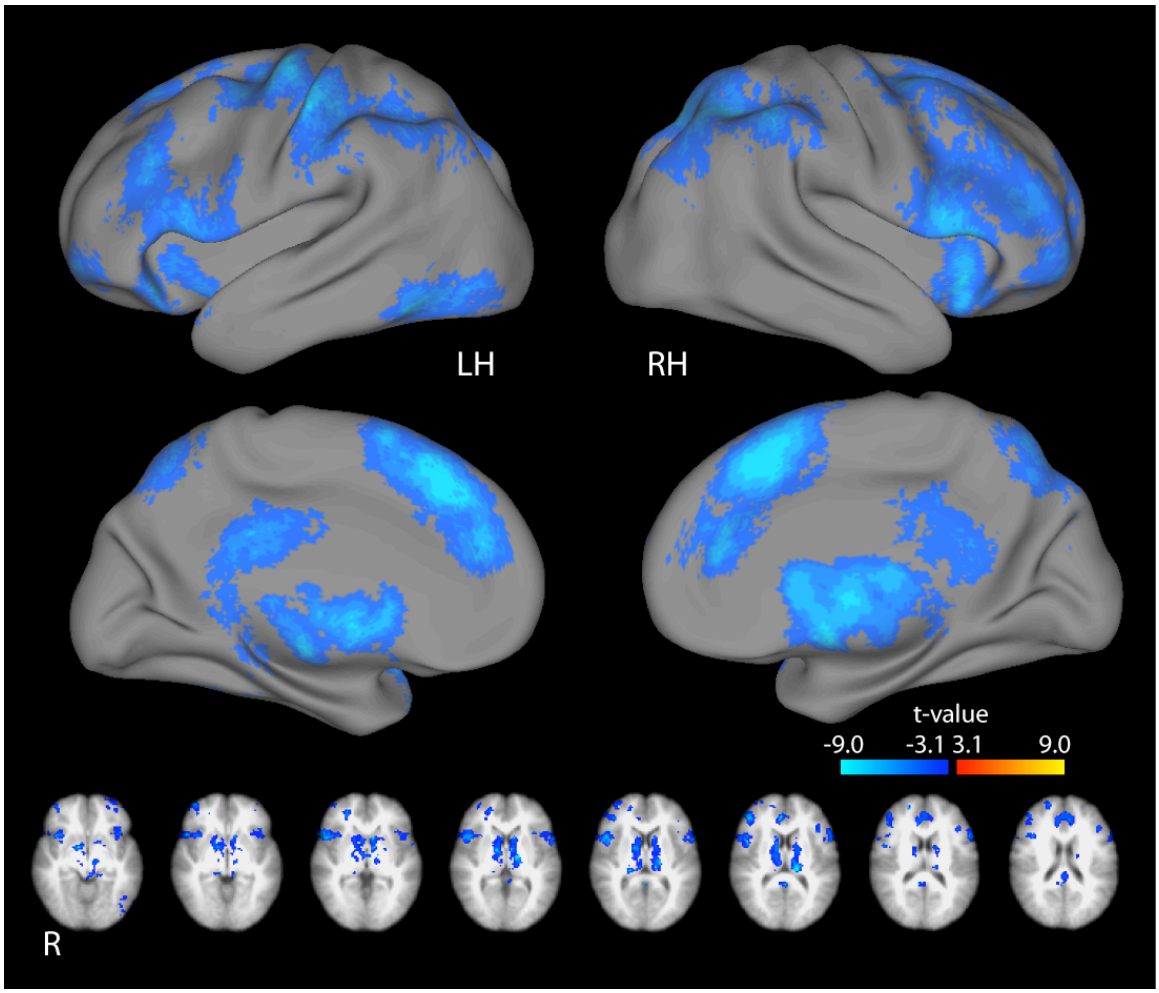

Figure S6. Regions showing significant positive correlation between ln(λ) and neural loss aversion (difference between slopes of neural loss and gain responses) (whole brain false discovery rate corrected at $q < 0.05$ [$t > 3.7$] and cluster extent > 100 voxels). No regions showed significant negative correlation.

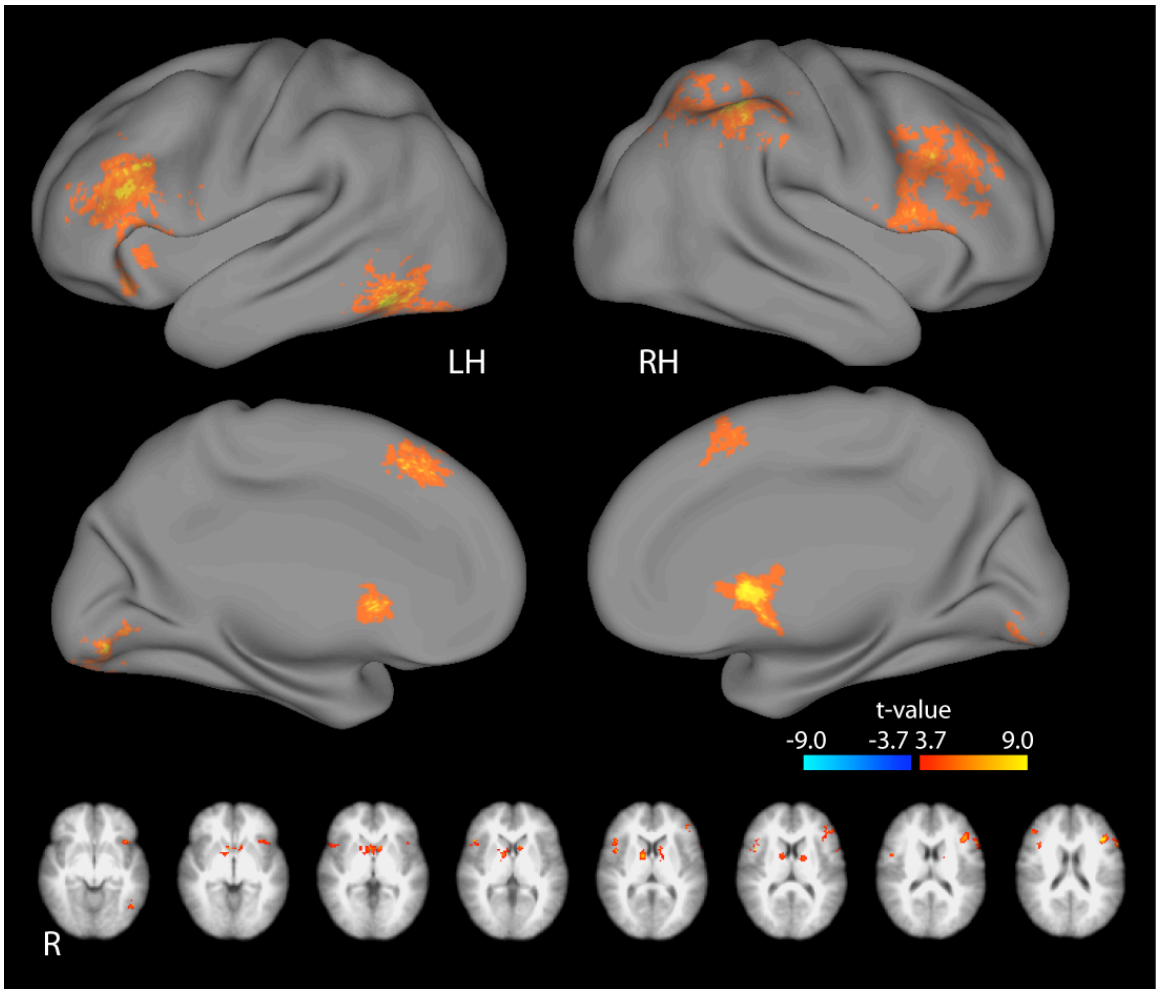

Figure S7. Correspondence between neural and behavioral loss aversion. Left panel presents statistical maps of the correlation between neural and behavioral loss aversion in whole brain analysis (whole brain false discovery rate corrected at $q < 0.05$ [$t > 3.7$] and cluster extent > 100 voxels) (see also Figure S6 and Table S2). Right panel presents scatterplots of behavioral versus neural loss aversion in several clusters. Regression lines and p-values were computed using robust regression by iteratively-reweighted least squares, to prevent influence of outliers. MNI coordinates (X/Y/Z center of gravity in mm) for plotted clusters: B ventral striatum (3.6, 6.3, 3.9), L inferior/middle frontal (-48.5, 24.7, 17.0), R inferior frontal (50.2, 14.3, 7.6), R inferior parietal (47.9, -45.6, 49.4). All correlations remained significant ($p < .05$) when the extreme individual was removed.

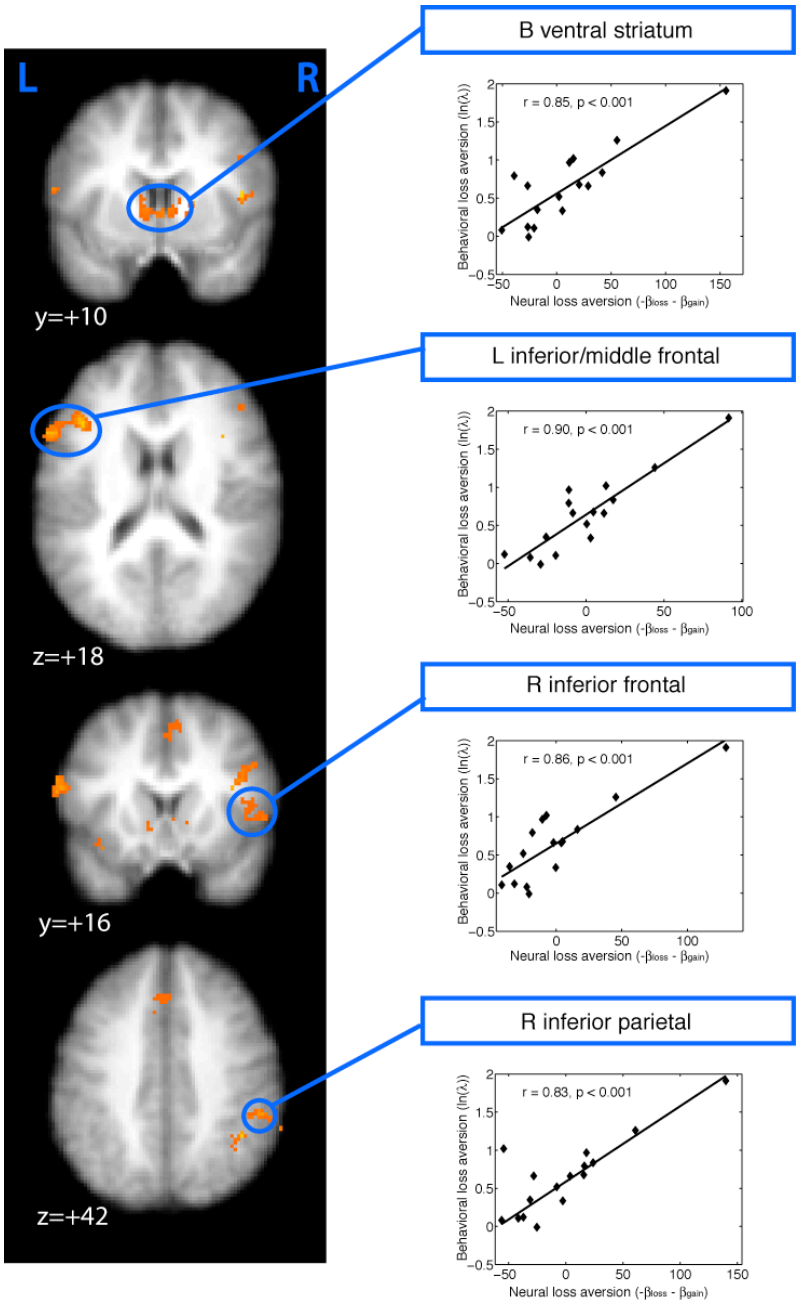

Table S1. Locations of significant activation in conjunction analysis for potential loss and gain effects ($Z > 2.3$, whole-brain corrected $p<.05$ in each map, extent in conjunction map ≥ 8 voxels). MNI coordinates denote the 3-dimensional center of gravity of each cluster. Mean Z statistics were created by averaging statistical maps over all voxels in cluster for parametric gain and loss analyses.

Location	**Cluster extent (voxels)**	**MNI X (mm)**	**MNI Y (mm)**	**MNI Z (mm)**	**Mean Z statistic (gains)**	**Mean Z statistic (losses)**
B striatum (nuc. Accubens, caudate), thalamus	1639	-0.4	6.1	-1.5	2.97	2.89
B VMPFC/OFC	1001	-6.0	39.3	-8.4	2.69	2.83
L frontal pole	154	-15.9	66.7	7.4	2.77	2.87
L middle frontal gyrus	116	-20.4	30.3	49.8	2.62	2.78
R middle/superior frontal gyrus	59	23.2	36.2	28.0	2.66	2.71
R frontal pole	48	7.0	63.9	18.1	2.58	2.65
R posterior cingulate	28	8.5	-38.3	32.5	2.66	2.66
R midbrain	8	10.2	-13.8	-16.0	2.59	2.68

Table S2. Locations of significant relation between ln(λ) and neural loss aversion (NLA) using robust regression (whole brain false discovery rate corrected at $q < 0.05$ [$t > 3.7$] and cluster extent > 100 voxels). Reported Pearson r-values were computed between ln(λ) and parametric gain and loss responses and NLA averaged across all voxels in each cluster; p-values for these correlations were computed using robust regression. Confidence intervals (CI) for indirect effect were estimated using the bias-corrected and accelerated bootstrap method described by Preacher & Hayes (2004). For all columns, asterisk denotes significant effects at $p < 0.05$.

Location	**Voxels**	**MNI X**	**MNI Y**	**MNI Z**	**r(ln(l), gain)**	**r(ln(l), loss)**	**r (ln(l), NLA)**	**Indirect effect CI**
L inferior/middle frontal	284	-48.5	24.7	17.0	0.11	-0.82*	0.9*	(0.0013, 0.0308) *
R inferior/middle frontal	175	47.5	22.4	26.0	0.28	-0.81*	0.88*	(0.0102, 0.0342) *
L inferior frontal (opercular)/ anterior insula	104	-39.5	19.8	-8.2	0.44	-0.82*	0.87*	(0.0070, 0.0213)*
R inferior frontal (opercular)	122	50.2	14.3	7.6	0.36	-0.91*	0.86*	(-0.0122, 0.0218)
B Ventral striatum	332	3.6	6.3	3.9	0.38	-0.85*	0.85*	(0.0071, 0.0279) *
R inferior parietal	358	47.9	-45.6	49.4	0.29	-0.87*	0.83*	(-0.0011, 0.0201)
B pre-SMA	110	-0.2	22.0	48.1	0.50*	-0.83*	0.81*	(0.0011, 0.0170)*
L lateral occipital/ cerebellum	963	-29.4	-74.3	-25.5	0.17	-0.55*	0.46*	(-0.0262, 0.0211)

References

1. R. H. Thaler, E. Johnson, *Management Science* **36**, 643 (1990).
2. A. M. Dale, *Hum Brain Mapp* **8**, 109 (1999).
3. A. Tversky, D. Kahneman, *Journal of Risk and Uncertainty* **5**, 297 (OCT, 1992).
4. M. Jenkinson, P. Bannister, M. Brady, S. Smith, *Neuroimage* **17**, 825 (Oct, 2002).
5. M. Jenkinson, S. Smith, *Med Image Anal* **5**, 143 (Jun, 2001).
6. M. W. Woolrich, B. D. Ripley, M. Brady, S. M. Smith, *Neuroimage* **14**, 1370 (Dec, 2001).
7. M. W. Woolrich, T. E. Behrens, C. F. Beckmann, M. Jenkinson, S. M. Smith, *Neuroimage* **21**, 1732 (Apr, 2004).
8. C. F. Beckmann, M. Jenkinson, S. M. Smith, *Neuroimage* **20**, 1052 (Oct, 2003).
9. K. J. Worsley, A. C. Evans, S. Marrett, P. Neelin, *J Cereb Blood Flow Metab* **12**, 900 (Nov, 1992).
10. T. E. Nichols, A. P. Holmes, *Hum Brain Mapp* **15**, 1 (Jan, 2002).
11. T. D. Wager, M. C. Keller, S. C. Lacey, J. Jonides, *Neuroimage* **26**, 99 (May 15, 2005).
12. C. R. Genovese, N. A. Lazar, T. Nichols, *Neuroimage* **15**, 870 (Apr, 2002).
13. D. C. Van Essen, *Neuroimage* **28**, 635 (Nov 15, 2005).
14. T. Nichols, M. Brett, J. Andersson, T. Wager, J. B. Poline, *Neuroimage* **25**, 653 (Apr 15, 2005).
15. K. J. Preacher, A. F. Hayes, *Behav Res Methods Instrum Comput* **36**, 717 (Nov, 2004).