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]

1. Department of Psychology, UCLA

2. Anderson School of Management, UCLA

3. Brain Research Institute, UCLA

Abstract

One of the most robust phenomena in behavioral studies of decision making is loss aversion, the tendency for people to exhibit greater sensitivity to losses than to equivalent sized gains. We measured brain activity while individuals decided whether to accept or reject gambles without feedback. This design isolated activity reflecting decisions without contamination by the anticipation or experience of impending monetary gains or losses. A broad neural network (including midbrain dopaminergic regions and their limbic and cortical targets) showed increasing activity as the potential gain increased, whereas an overlapping set of regions showed decreasing activity as the potential loss increased. Thus, potential losses did not engage a separate set of emotional brain systems, but were instead represented by decreasing activity in several gain-sensitive areas. Moreover, these regions exhibited neural loss aversion as shown by their

greater sensitivity to losses than gains. Finally, individual differences in *behavioral* loss aversion were predicted by a measure of *neural* loss aversion in several regions including ventral striatum and prefrontal cortex. These results provide the first neuroscientific evidence that risk aversion is driven by the brain's greater sensitivity to losses than gains.

Main text

Many decisions, such as whether to invest in the stock market or whether to accept a new job, involve the possibility of either gaining or losing relative to the status quo. When faced with such decisions, most people are strikingly risk averse. For instance, when deciding whether to accept gambles that offer a 50/50 chance of gaining or losing money, people typically only accept gambles in which the amount that could be gained is at least twice the amount that could be lost (e.g., a 50/50 chance to either gain $100 or lose $50) (*1*). Prospect theory, the most successful behavioral model of decision making under risk and uncertainty (*1, 2*), explains risk aversion for "mixed" (gain/loss) gambles using the concept of *loss aversion*: People are more sensitive to the possibility of losing objects or money than they are to the possibility of gaining the same objects or amounts of money (*1, 3, 4*). Thus, people typically require a potential gain of at least $100 to make up for exposure to a potential loss of $50 because the subjective impact of losses is roughly twice that of gains.

Loss aversion also applies to decisions that involve no risk (*5*). For example, in one study, students who were randomly assigned to receive a coffee mug (so that giving up the mug would be perceived as a loss) subsequently set a median selling price of $7.12 for that mug, whereas students who were randomly assigned an opportunity to receive either an identical mug or money (so that acquiring the mug would be perceived as a

gain) valued the mug at only $3.12 (*6*). Outside the laboratory, loss aversion has been invoked to help explain a wide range of economic behaviors, such as why consumer demand for products is more sensitive to price increases than decreases (*7*) and why investors require a much higher average return to invest in stocks than bonds (*8*). Moreover, loss aversion has been documented in the trading behavior of children as young as age five (*9*) and capuchin monkeys (*10*), suggesting that the tendency for losses to loom larger than gains may reflect a fundamental feature of how potential outcomes are assessed by the primate brain.

The neural basis of loss aversion has not been directly investigated to date. Previous neuroimaging studies of responses to monetary gains or losses have focused on activity associated with the anticipation of immediate outcomes ("anticipated" utility) (*11, 12*) or the actual experience of gaining or losing money ("experienced" utility) (*11, 13, 14*). However, to our knowledge, no previous studies have addressed the question of which brain systems represent potential losses versus gains when a decision is being made ("decision" utility). Behavioral researchers have shown that evaluations of outcomes at these different points in time (i.e., "decision", "anticipated", and "experienced", utilities) often diverge in dramatic ways, which raises the possibility that the corresponding brain systems involved may also differ (*15*). In the current study we aimed to isolate activity associated with the evaluation of a gamble when choosing whether or not to accept it (i.e., "decision" utility). This allowed us to test whether the neural response during the evaluation of potential outcomes is similar to patterns previously reported in studies of anticipated and experienced outcomes.

One important question is whether loss aversion reflects cognitive or emotional processes. Decisions have been shown to vary systematically depending on whether options are described in terms of gains from one reference point or losses from another (*16*), demonstrating that cognitive representations are involved in loss aversion. However, this cognitive account alone does not explain why losses should loom larger than gains. Alternatively, it has been suggested that enhanced sensitivity to losses is driven by negative emotional responses, such as fear or anxiety (e.g, *17*). This notion predicts that exposure to increasing potential losses should be associated with increased activity in brain systems involved in negative emotions (such as amygdala or anterior insula, cf., (*18, 19*)). Alternatively, loss aversion could reflect an asymmetric response to losses versus gains within a single system that codes for the subjective value of the potential gamble, such as ventromedial/orbital prefrontal cortex and ventral striatum (*11, 20, 21*).

Imaging decision utility. To examine the neural systems that process decision utility, we collected functional magnetic resonance imaging (fMRI) data while participants decided whether to accept or reject "mixed" gambles that offer a 50/50 chance of either gaining one amount of money or losing another amount (see Figure 1) (*22*). To encourage participants to reflect on the subjective attractiveness of each gamble rather than revert to a fixed decision rule, we asked that they indicate one of four responses to each gamble (strongly accept, weakly accept, weakly reject, and strongly reject). In order to estimate the neural response to gains and losses separately, the sizes of the potential gain and loss were manipulated independently, with gains ranging from \$10 to \$40 (in increments of \$2) and losses ranging from \$5 to \$20 (in increments of \$1).

These ranges were chosen because previous studies indicate that people are, on average, roughly twice as sensitive to losses as to gains (*1, 23*); thus, we expected that for most participants this range of gambles would elicit the entire range of attitudes, from strong acceptance to indifference to strong rejection. All combinations of gains and losses were presented (see Figure 1), with trials distributed over three fMRI scanning runs. During scanning, participants evaluated each gamble without receiving feedback. In addition to allowing for the isolation of decision utility from anticipated or experienced utility, it prevented early trial outcomes from influencing a participant's decision making on later trials.

Sixteen participants were endowed with $30 at least one week prior to the scanning session. The endowment was provided during a separate behavioral testing session in order to minimize the potential increase in risk seeking behavior that can occur when individuals feel that they are gambling with "house money" (*24*). Participants were asked to bring $60 in cash to the scanning session, and told they could win up to an additional $120 or lose up to $60. At the outset of the scanning session, participants were told that one of their decisions from each of the three scanning runs would be selected at random, and each selected gamble would be played for real money if they had accepted it during scanning. Outcomes were summed across all "played" trials and ranged from -$12 to +$70, with an average total gamble outcome of $23 (leading to an average payout of $53, including the endowment).

---- Figure 1 about here ----

Behavioral loss aversion. Behavioral sensitivity to gains and losses was assessed by fitting a logistic regression to each participant's acceptability judgments collected

during scanning, using the size of the gain and loss as independent variables. Based on this analysis, a measure of loss aversion (λ) was computed as the ratio of the (absolute) loss response to the gain response. The observed level of loss aversion (median $\lambda = 1.93$; range: 0.99-6.75) is consistent with the fact, shown in Figure 1B, that participants were roughly indifferent to gambles in which the potential gain was twice the potential loss (i.e., gambles that fall along the main diagonal of the gain/loss matrix in Figure 1). This finding also accords well with results reported by other researchers (*1*, *23*).

Neural response to potential gains and losses. The imaging data were first analyzed to identify regions whose activation correlated with the size of the potential gain or loss, using parametric regressors (for details on model specification, see Supplementary Materials and Methods). This analysis found a network of regions responsive to the size of potential gains when evaluating gambles (averaging over levels of loss) (see Figure 2). The gain-responsive network included regions previously shown to be associated with the anticipation and receipt of monetary rewards, including dorsal and ventral striatum, ventromedial prefrontal cortex (vmPFC), ventrolateral PFC (vlPFC), anterior cingulate (ACC), orbitofrontal cortex (OFC), and dopaminergic midbrain regions (see Supplementary Figure 2). There were no regions that showed decreasing activation as gains increased.

---- Figures 2 and 3 about here ----

If loss aversion is driven by a negative affective response (e.g., fear, vigilance, discomfort), then one would expect increasing activity in brain regions associated with these emotions as the size of the potential loss increases. Contrary to this prediction, no brain regions showed significantly increasing activation during evaluation of gambles as

the size of the potential loss increased (averaging over all levels of gain). Instead, a network of regions including the striatum, vmPFC, ventral ACC, and medial OFC, most of which also coded for gains, showed *decreasing* activity as the size of the potential loss increased (see Figure 2 and Supplementary Figure 3). A conjunction analysis between increasing activity for gains and decreasing activity for losses demonstrated joint sensitivity to both gains and losses in a set of regions, including the dorsal and ventral striatum and vmPFC (see Figure 3 and Supplementary Table 1).

In order to ensure that potential loss-related responses were not being obscured by the overall positive expected value of the gambles, we compared activity evoked by the worst possible gambles (gain $10-$16, loss $17-$20) and the best possible gambles (gain $34-$40, loss $5-8). In a whole-brain analysis, there were no regions that showed significantly more activity for the worst gambles compared to the best (corrected $p > 0.4$ in all voxels using randomization tests). Given the specific prediction regarding loss-related activity in amygdala and insula based on prior studies of experienced utility and risk aversion (*11, 19*), we performed further analyses focused on these areas. Even at a very liberal uncorrected threshold of $p < 0.01$, there were no significant voxels in the amygdala, and only two single unconnected voxels in the insula. By comparison, at the same threshold there were large clusters of activation for the best versus the worst gambles in the ventral striatum and vmPFC. These results strongly support the conclusion that losses and gains are coded by the same mechanism rather than two separate mechanisms. Moreover, this aggregate representation of decision utility appears to be represented by the same neural circuitry that is engaged by experienced rewards such as the receipt of money (*11*), the consumption of cocaine (*25*) or fruit juice (*26*), and

the viewing of attractive faces (*27*). These results are also consistent with previous studies showing increased and decreased activity in striatum, respectively, for experienced monetary gains and losses (*11, 13*).

Correlating behavioral and neural loss aversion. As noted above, there was substantial variability in behavioral loss aversion (i.e., reluctance to gamble) across individuals in the present study. We next investigated whether individual differences in brain activity during decision making were related to these individual differences in behavioral loss aversion, using whole-brain analyses to identify regions where the neural response to gains or losses was correlated with behavioral loss aversion. Surprisingly, greater loss aversion was associated with greater sensitivity to both not only losses but also gains. For increasing gains, correlation with behavioral loss aversion was only observed in the sensorimotor cortex and superior frontal cortex (see Supplementary Figure 4). On the other hand, as potential losses increased, an extensive set of areas showed a more rapidly decreasing response to mounting losses in individuals who were more loss averse (see Supplementary Figure 5). Notably, these regions encompassed many of the areas that showed an overall decrease in neural activity with increasing potential loss. The association of decreased loss aversion with decreased neural responses to losses and gains during decision making is consistent with the longstanding notion that some forms of risk-taking may have their roots in sensation-seeking by individuals who have a diminished physiological response to stimulation (*28*). It is also worth remarking that previous work has shown that individuals with high levels of the "harm avoidance" personality trait (who are presumably more loss averse) show greater activation of the ventral striatum during risky decision making (cf., *29*).

Initial examination of regions of interest in the striatum and vmPFC from the gain/loss conjunction analysis (as shown in Figure 3) revealed that these regions exhibit a pattern of "neural loss aversion"; that is, the (negative) slope of the decrease in activity to increasing losses was steeper than the slope of the increase in activity for increasing gains in a majority of participants (striatum: median loss/gain =1.88, loss > gain for 14/16 participants, sign test $p = .004$; vmPFC: median loss/gain = 1.82, loss > gain for 13/16 participants, $p = .021$). Therefore, whereas loss averse participants showed more neural sensitivity to both potential gains and losses compared to less loss averse participants, it appears that this neural sensitivity was disproportionately greater for losses than for gains.

In order to more directly assess the relationship between behavioral loss aversion and neural loss aversion, we defined a neural loss aversion measure as the difference between the (negative) slope of the parametric loss responses and the slope of the parametric gain responses at each voxel. Whole-brain analysis of the correlation between neural and behavioral loss aversion (see Supplementary Figure 6 and Supplementary Table 2) showed significant correlations in several regions, including bilateral lateral and superior (pre-SMA) PFC, bilateral ventral striatum, right inferior parietal cortex, and lateral occipital/cerebellum. Activation maps and scatterplots for a subset of these regions are shown in Figure 4. These analyses show exceedingly strong correlations between behavioral and neural loss aversion, and these associations are highly significant using robust regression, which prevents undue influence of outliers.

---- Figure 4 about here ----

Loss aversion is mediated by sensitivity to losses. 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 Supplementary Table 2). 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). Thus, differences in behavioral loss aversion across individuals seem to be driven primarily by the degree to which those individuals show decreasing neural activity to potential losses.

Conclusions. The present study replicates the common behavioral pattern of risk aversion for mixed gambles that offer a 50/50 chance of gaining or losing money, and shows for the first time that this pattern of behavior is directly tied to the brain's relative sensitivity to potential losses versus gains. These results provide striking evidence in favor of one of the fundamental claims of prospect theory (*1, 2*), namely that the function mapping money to subjective value is steeper for losses than gains. Importantly,

mediation analysis suggests that individual differences in risk attitudes (as measured by the behavioral coefficient of loss aversion) are driven primarily by individual differences in the degree to which activity in the brain's reward circuitry is attenuated by potential losses. Although the present study focused on loss aversion in the context of mixed gambles, recent work has found that the coefficient of loss aversion (i.e., the ratio of sensitivity to losses versus gains) is highly correlated across risky and riskless contexts (*30*). Therefore, we surmise that a similar mechanism may contribute to other manifestations of loss aversion.

Neural loss aversion was observed throughout, though not strictly limited to, the targets of the mesolimbic and mesocortical dopamine (DA) systems (specifically, the ventral striatum and orbital, medial, and lateral prefrontal cortices). Lesions to some of these regions (particularly the core region of nucleus accumbens in the ventral striatum and medial PFC) have been previously associated with increased risk aversion in animal models (*31*), and disruption of the DA system can also result in shifts of risk attitudes in humans (*32*). It is tempting to speculate that the individual differences in behavioral and neural loss aversion observed in the present study may be related to naturally occurring differences in dopamine function, perhaps due to genetic variability in dopamine signaling or metabolism. This is an appealing hypothesis, though the relation between genetic variation in the DA system and personality traits such as impulsivity and risk-taking remains largely unknown (*33*), and it is possible that other neurotransmitter systems (e.g., serotonin and noradrenaline) are also involved.

Previous studies have shown that anticipated or experienced losses give rise to activation in regions that have been associated with negative emotions, such as amygdala

or anterior insula (*11, 18, 19*). In contrast, the present study demonstrates that, in the context of decision making, potential losses are represented by decreasing activity in regions that seem to code for subjective value rather than by increasing activity in regions coding for negative emotions. This difference between present and prior results reinforces the importance of distinguishing between experienced, anticipated, and decision utility in economic theories of choice (*15*). It is possible that the engagement of the amygdala for experienced losses reflects negative prediction error (*11, 34*) rather than negative value per se, whereas the lack of immediate outcomes in the present study precludes the computation of prediction errors.

The present results also illustrate how neuroimaging can be used to directly test predictions stemming from behavioral theories, and thereby complement evidence provided by behavioral studies and studies of lesion patients. Further, the diminished neural sensitivity to losses among less loss averse (i.e., more risk seeking) individuals may shed light on a number of neuropsychiatric and behavioral disorders, such as substance abuse, pathological gambling, and antisocial personality disorder, that are associated with increased risk-taking and impulsive behavior. The involvement of dopamine in these disorders, and the overlap of our results with the dopamine system and its targets, suggest that greater traction on these disorders could be obtained by integrating the methods and models of behavioral economics with the tools of cognitive neuroscience.

Acknowledgments

This work was supported by NSF DMI-0433693 (C. Fox & R. Poldrack, PIs), and by NIH P20 RR020750 (R. Bilder, PI). The authors thank Adam Aron, Robert Bilder,

Michael Frank, Adriana Galvan, Marisa Geohegan, Eric Johnson, Matthew Lieberman, Raj Raizada, and Elena Stover for helpful comments, Kristopher Preacher for assistance with mediation analysis, and Ajay Satpute and Jessica Cohen for assistance in data collection.

Figure Captions

Figure 1. A. An illustration of the event-related task design. During each trial, the participant was presented for 3 seconds with a display showing the size of the potential gain (in green) and loss (in red). After the acceptability response, a variable interval was presented to allow for optimal deconvolution of fMRI responses to each trial (*35*). Gambles were not resolved during scanning. The values of gain and loss for each trial were sampled from the gain/loss matrix, as shown here for two example gambles; a gamble from each cell in this 16×16 matrix was presented during scanning, but the data were collapsed into a 4×4 matrix for analysis. B. Color-coded heatmap of probability of gamble acceptance at each level of gain/loss (red: high willingness to accept the gamble, blue: low willingness to accept the gamble). Participants' willingness to accept gambles increased as the size of the gain increased, and decreased as the size of the loss increased. C. Color-coded heatmap of response times (red: slower response times, blue: faster response times). Performance was slowest on trials that were closest to the point of indifference between acceptance and rejection.

Figure 2. Whole-brain analysis of parametric responses to size of potential gain (left) or loss (right). Statistical maps were projected onto an average cortical surface using multifiducial mapping in CARET (*36*); coronal slices (y=+10) are included to show ventral striatal activation. All maps are corrected for multiple comparisons at the whole-brain level using cluster-based Gaussian random field correction (*37*) at $p < .05$.

Figure 3. Conjunction analysis results. Map in left panel shows regions with conjointly significant positive gain response and negative loss response ($p < .05$, whole-brain corrected, in each individual map). Heatmaps on right were created by averaging

parameter estimates versus baseline within each cluster in the conjunction map for each of the 16 cells (of 16 gambles each) in the gain/loss matrix; color-coding reflects strength of neural response for each condition, such that dark red represents the strongest activation and dark blue represents the strongest deactivation.

Figure 4. 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 Supplementary Figure 6 and Supplementary Table 2). 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).

References

1. A. Tversky, D. Kahneman, *Journal of Risk and Uncertainty* **5**, 297 (OCT, 1992).
2. D. Kahneman, A. Tversky, *Econometrica* **4**, 263 (1979).
3. N. Novemsky, D. Kahneman, *Journal of Marketing Research* **42**, 119 (2005).
4. M. Rabin, *Econometrica* **68**, 1281 (2000).
5. A. Tversky, D. Kahneman, *Quarterly Journal of Economics* **106**, 1039 (NOV, 1991).
6. D. Kahneman, J. Knetsch, R. H. Thaler, *Journal of Political Economy* **98**, 1325 (1990).
7. B. Hardie, E. Johnson, P. Fader, *Marketing Science* **12**, 378 (1993).
8. S. Benartzi, R. H. Thaler, *Quarterly Journal of Economics* **110**, 73 (FEB, 1995).
9. W. T. Harbaugh, K. Krause, L. Vesterlund, *Economic Letters* **70**, 175 (2001).
10. M. K. Chen, V. Lakshminarayanan, L. R. Santos, *Journal of Political Economy* (in press).
11. H. C. Breiter, I. Aharon, D. Kahneman, A. Dale, P. Shizgal, *Neuron* **30**, 619 (May, 2001).
12. B. Knutson, C. M. Adams, G. W. Fong, D. Hommer, *J Neurosci* **21**, RC159 (Aug 15, 2001).
13. M. R. Delgado, H. M. Locke, V. A. Stenger, J. A. Fiez, *Cogn Affect Behav Neurosci* **3**, 27 (Mar, 2003).
14. B. Knutson, G. W. Fong, C. M. Adams, J. L. Varner, D. Hommer, *Neuroreport* **12**, 3683 (Dec 4, 2001).

15. D. Kahneman, P. P. Wakker, R. Sarin, *Quarterly Journal of Economics* **112**, 375 (MAY, 1997).

16. A. Tversky, D. Kahneman, *Science* **211**, 453 (Jan 30, 1981).

17. C. F. Camerer, *Journal of Marketing Research* **42**, 129 (2005).

18. I. Kahn *et al.*, *Neuron* **33**, 983 (Mar 14, 2002).

19. C. M. Kuhnen, B. Knutson, *Neuron* **47**, 763 (Sep 1, 2005).

20. B. Knutson, G. W. Fong, S. M. Bennett, C. M. Adams, D. Hommer, *Neuroimage* **18**, 263 (Feb, 2003).

21. S. M. McClure *et al.*, *Neuron* **44**, 379 (Oct 14, 2004).

22. . (Materials and methods are available as supporting materials on Science Online.).

23. M. Abdellaoui, H. Bleichrodt, C. Paraschiv. (ENSAM-Paris, France, 2005).

24. R. H. Thaler, E. Johnson, *Management Science* **36**, 643 (1990).

25. H. C. Breiter *et al.*, *Neuron* **19**, 591 (Sep, 1997).

26. G. Pagnoni, C. F. Zink, P. R. Montague, G. S. Berns, *Nat Neurosci* **5**, 97 (Feb, 2002).

27. I. Aharon *et al.*, *Neuron* **32**, 537 (Nov 8, 2001).

28. M. Zuckerman, D. M. Kuhlman, *J Pers* **68**, 999 (Dec, 2000).

29. S. C. Matthews, A. N. Simmons, S. D. Lane, M. P. Paulus, *Neuroreport* **15**, 2123 (Sep 15, 2004).

30. E. Johnson, S. Gachter, A. Herrman. (Columbia Business School, 2006).

31. R. N. Cardinal, D. R. Pennicott, C. L. Sugathapala, T. W. Robbins, B. J. Everitt, *Science* **292**, 2499 (Jun 29, 2001).

32. R. Cools, R. A. Barker, B. J. Sahakian, T. W. Robbins, *Neuropsychologia* **41**, 1431 (2003).

33. E. Congdon, T. Canli, *Behav Cogn Neurosci Rev* **4**, 262 (Dec, 2005).

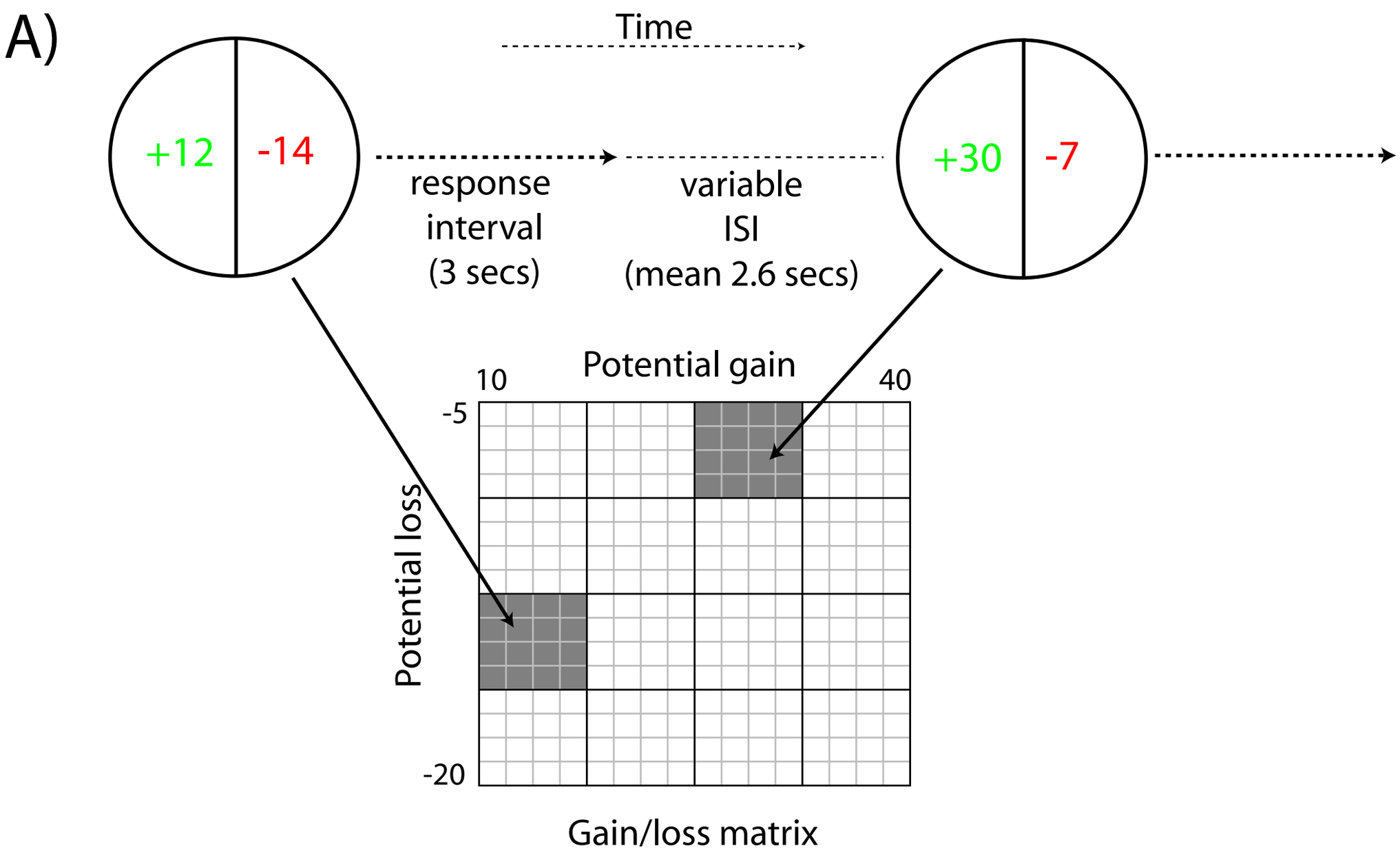
34. C. Büchel, in *Neuroimaging and psychological theories of memory*. (Marburg, Germany, 2006).

35. A. M. Dale, *Hum Brain Mapp* **8**, 109 (1999).

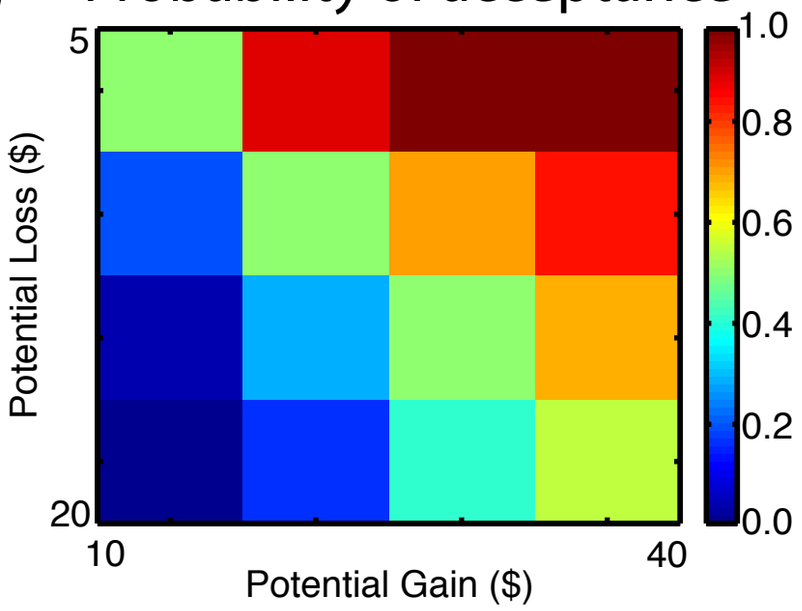
36. D. C. Van Essen, *Neuroimage* **28**, 635 (Nov 15, 2005).

37. K. J. Worsley, A. C. Evans, S. Marrett, P. Neelin, *J Cereb Blood Flow Metab* **12**, 900 (Nov, 1992).

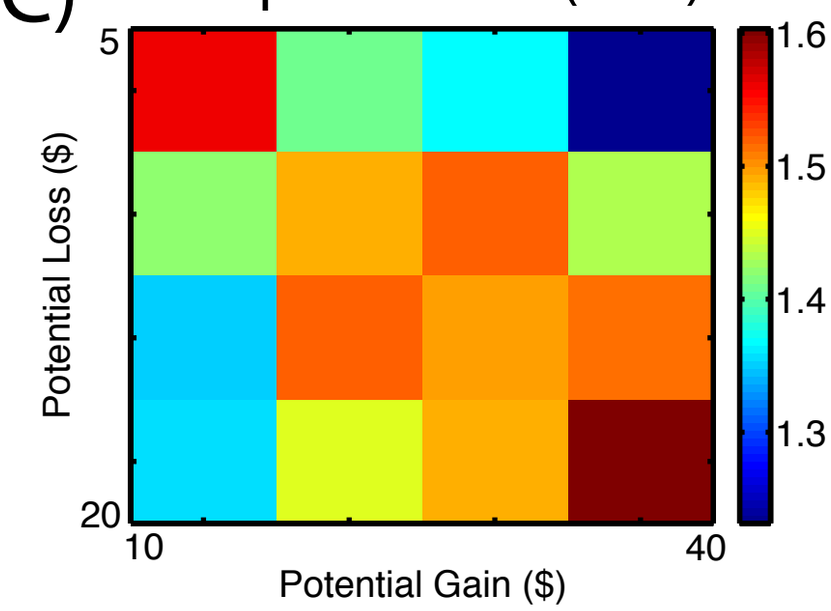
A)

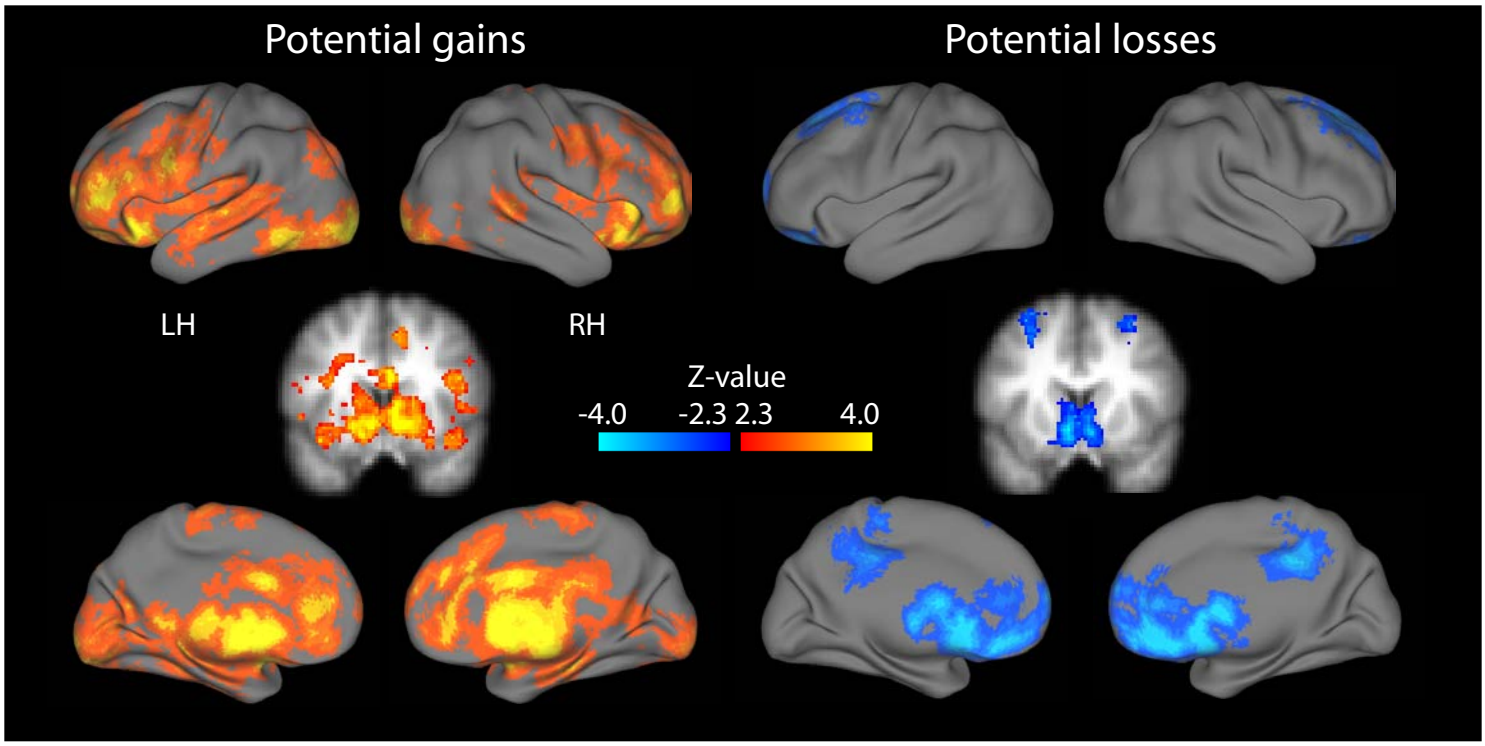


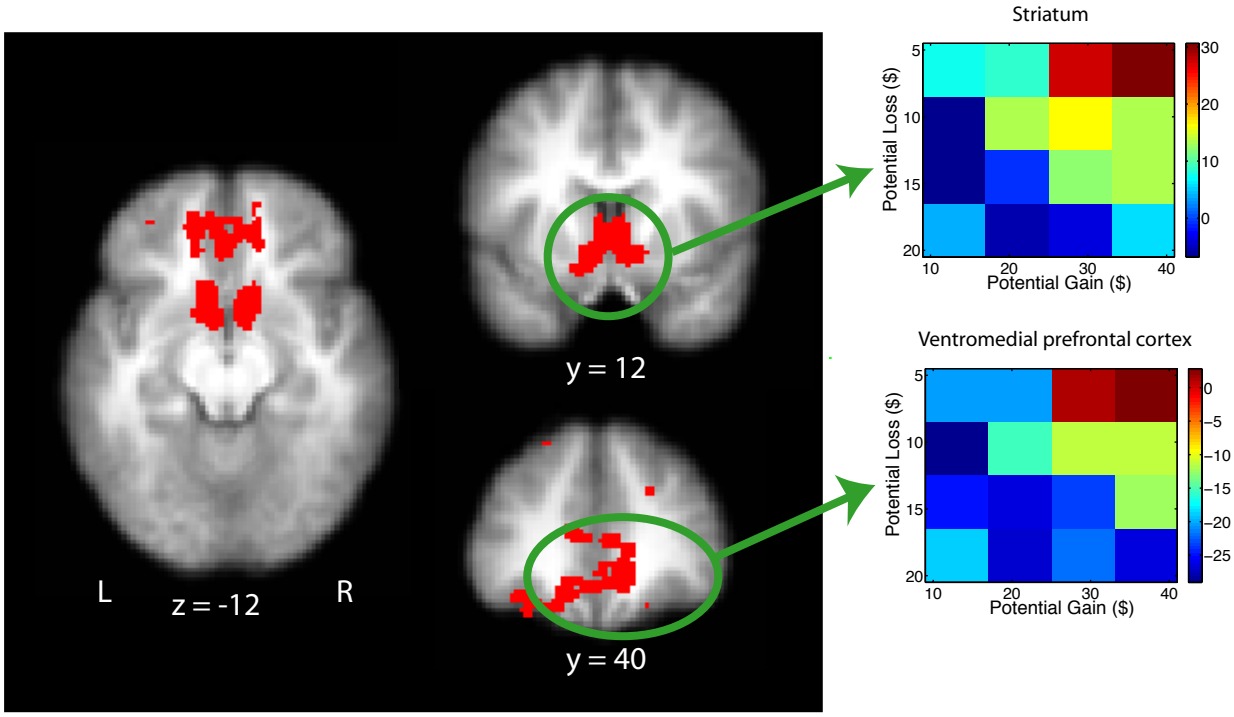
B) Probability of acceptance

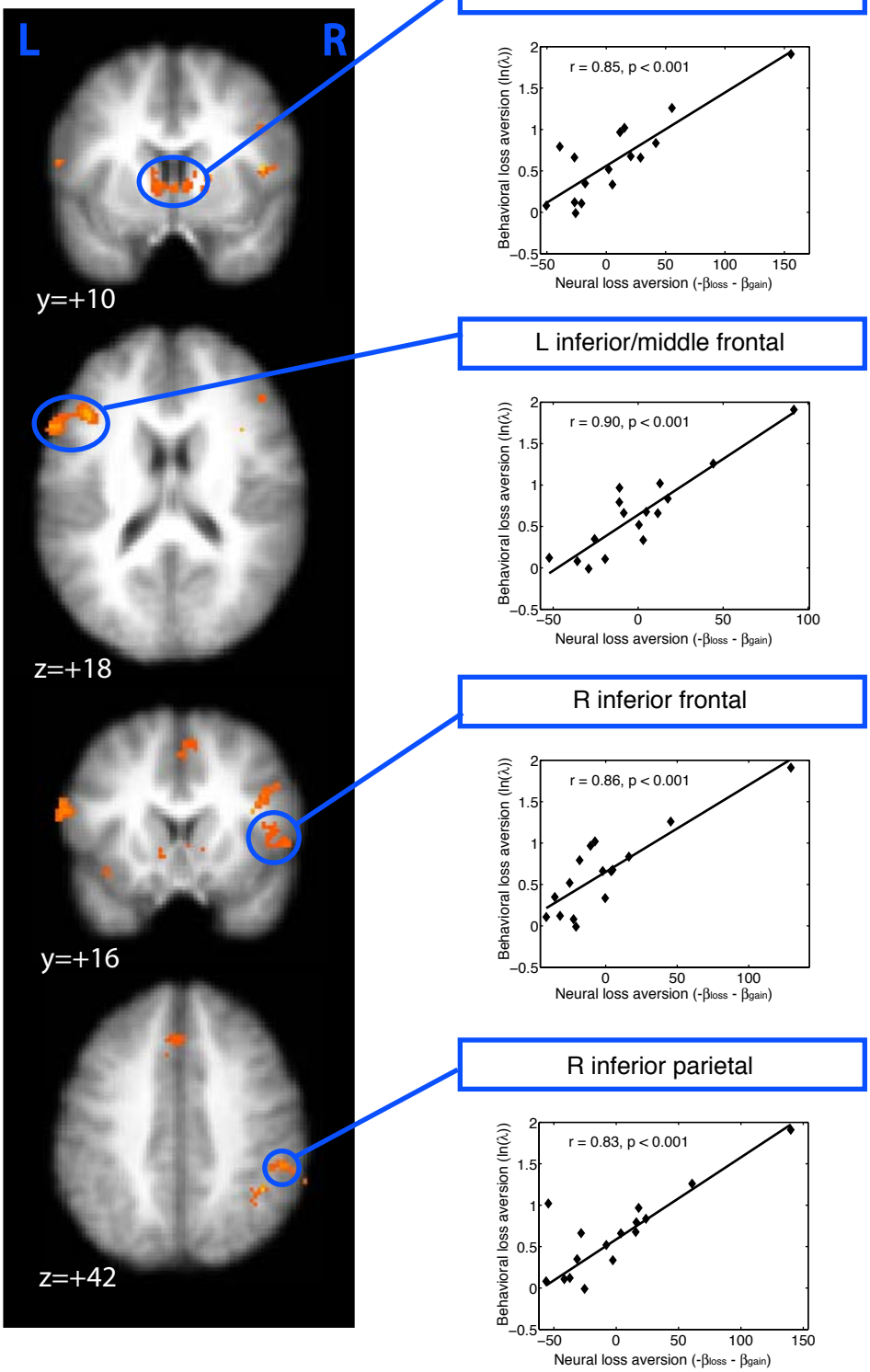


C) Response time (secs)




Potential gains
Potential losses
LH
RH
Z-value
-4.0
-2.3
2.3
4.0


L
z = -12
R
y = 12
y = 40
Striatum
Potential Loss ($)
Potential Gain ($)
5
10
15
20
10
20
30
40
30
20
10
0
Ventromedial prefrontal cortex
Potential Loss ($)
Potential Gain ($)
0
−5
−10
−15
−20
−25


L
R
y=+10
z=+18
y=+16
z=+42
B ventral striatum
r = 0.85, p < 0.001
Behavioral loss aversion (ln(λ))
Neural loss aversion (-βloss - βgain)
L inferior/middle frontal
r = 0.90, p < 0.001
Behavioral loss aversion (ln(λ))
Neural loss aversion (-βloss - βgain)
R inferior frontal
r = 0.86, p < 0.001
Behavioral loss aversion (ln(λ))
Neural loss aversion (-βloss - βgain)
R inferior parietal
r = 0.83, p < 0.001
Behavioral loss aversion (ln(λ))
Neural loss aversion (-βloss - βgain)

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]

1. **Department of Psychology, UCLA**
2. **Anderson School of Management, UCLA**
3. **Brain Research Institute, UCLA**

Supplementary 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. Thaler and Johnson, 1990). 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/) (Dale, 1999).

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 (Tversky and Kahneman, 1992) 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 (Jenkinson et al., 2002). 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 (Jenkinson and Smith, 2001). 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 Supplementary Figure 1), and three additional orthogonal parametric regressors were included representing: (a) the size of the potential gain (see Supplementary Figure 2), (b) the size of the potential loss (see Supplementary Figure 3), 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 (Woolrich et al., 2001) 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 (Beckmann et al., 2003; Woolrich et al., 2004), 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 (Worsley et al., 1992). For the comparison between best versus worst gamble conditions, control for multiple comparisions was implemented using randomization tests (Nichols and Holmes, 2002) 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 (see Supplementary Figures 4-6), voxelwise robust regression was used to in order to reduce the influence of outliers on the analysis (cf. Wager et al., 2005). 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 $q < .05$ (Genovese et al., 2002) 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 (2005). For the purposes of presentation, data are overlaid on the average atlas surface.

Conjunction analysis. Conjunction analysis for gains and losses (see Figure 2 in the main text and Supplementary Table 1) 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 $p < .05$ (Nichols et al., 2005).

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 (see Figure 6 and Table 2).

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 a data from each of the significant clusters in the neural loss aversion analysis (listed in Supplementary Table 2) as described by Preacher and Hayes (2004). 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.

References

Beckmann, C. F., Jenkinson, M., and Smith, S. M. (2003). General multilevel linear modeling for group analysis in FMRI. Neuroimage *20*, 1052-1063.

Dale, A. M. (1999). Optimal experimental design for event-related fMRI. Hum Brain Mapp *8*, 109-114.

Genovese, C. R., Lazar, N. A., and Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. Neuroimage *15*, 870-878.

Jenkinson, M., Bannister, P., Brady, M., and Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage *17*, 825-841.

Jenkinson, M., and Smith, S. (2001). A global optimisation method for robust affine registration of brain images. Med Image Anal *5*, 143-156.

Nichols, T., Brett, M., Andersson, J., Wager, T., and Poline, J. B. (2005). Valid conjunction inference with the minimum statistic. Neuroimage *25*, 653-660.

Nichols, T. E., and Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum Brain Mapp *15*, 1-25.

Preacher, K. J., and Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. Behav Res Methods Instrum Comput *36*, 717-731.

Thaler, R. H., and Johnson, E. (1990). Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes in Risky Choice. Management Science *36*, 643-660.

Tversky, A., and Kahneman, D. (1992). Advances in Prospect-Theory - Cumulative Representation of Uncertainty. Journal of Risk and Uncertainty *5*, 297-323.

Van Essen, D. C. (2005). A Population-Average, Landmark- and Surface-based (PALS) atlas of human cerebral cortex. Neuroimage *28*, 635-662.

Wager, T. D., Keller, M. C., Lacey, S. C., and Jonides, J. (2005). Increased sensitivity in neuroimaging analyses using robust regression. Neuroimage *26*, 99-113.

Woolrich, M. W., Behrens, T. E., Beckmann, C. F., Jenkinson, M., and Smith, S. M. (2004). Multilevel linear modelling for FMRI group analysis using Bayesian inference. Neuroimage *21*, 1732-1747.

Woolrich, M. W., Ripley, B. D., Brady, M., and Smith, S. M. (2001). Temporal autocorrelation in univariate linear modeling of FMRI data. Neuroimage *14*, 1370-1386.
Worsley, K. J., Evans, A. C., Marrett, S., and Neelin, P. (1992). A three-dimensional statistical analysis for CBF activation studies in human brain. J Cereb Blood Flow Metab *12*, 900-918.

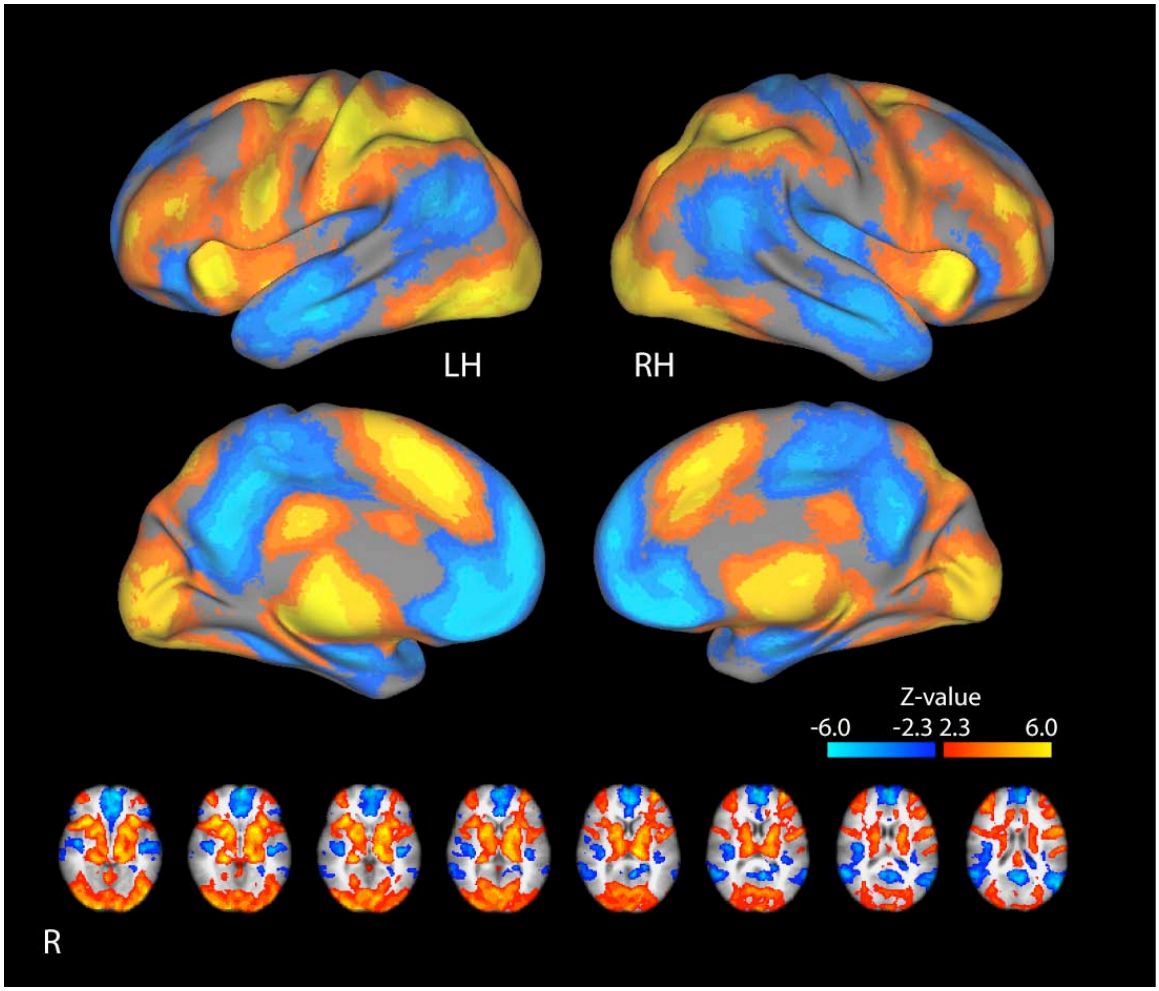
Supplementary Table 1. Locations of significant activation in conjunction analysis for potential loss and gain effects ($Z > 2.3$, whole-brain corrected $p<.05$ in each map, extent in conjunction map ≥ 8 voxels). MNI coordinates denote the 3-dimensional center of gravity of each cluster. Mean Z statistics 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

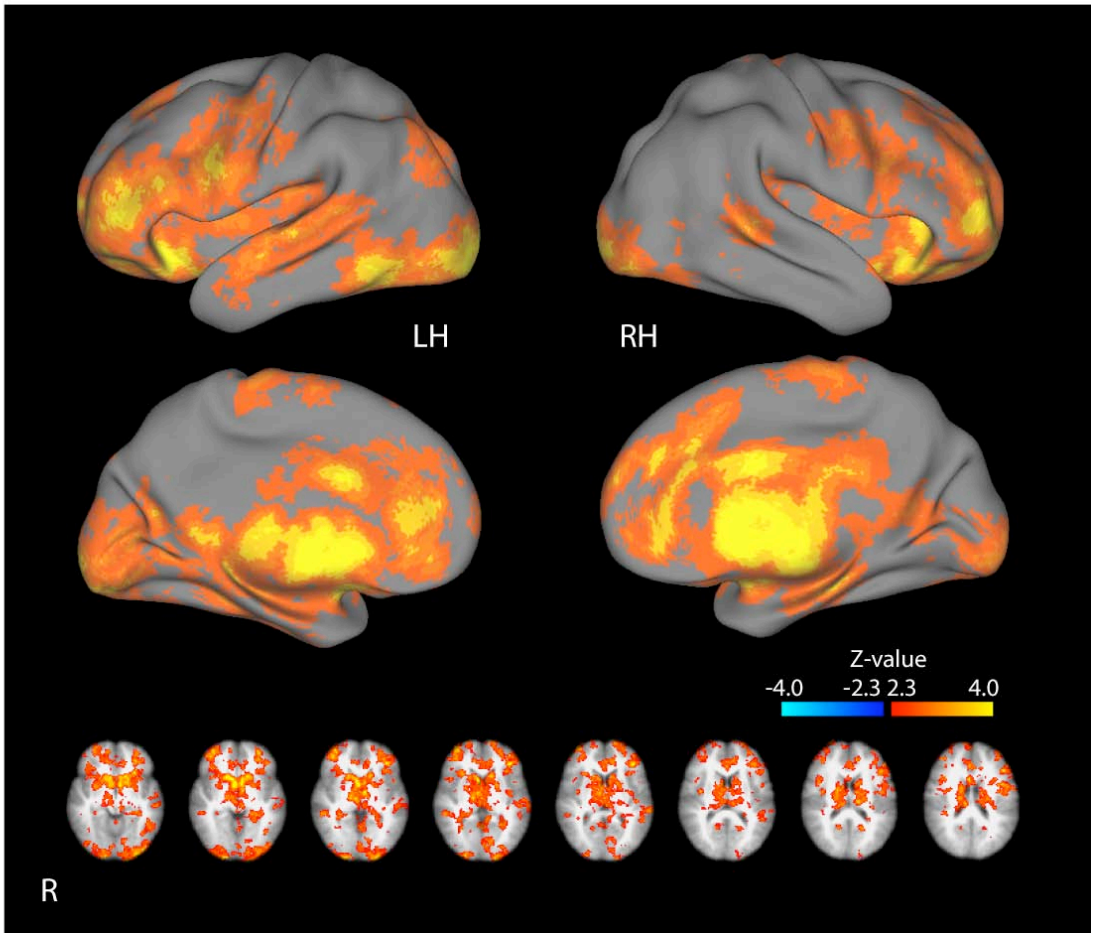
Supplementary Table 2. Locations of significant relation between ln(λ) and neural loss aversion (NLA) using robust regression (whole brain false discovery rate corrected at $q < 0.05$ [$t > 3.7$] and cluster extent > 100 voxels). Reported Pearson r-values were computed between ln(λ) and parametric gain 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)

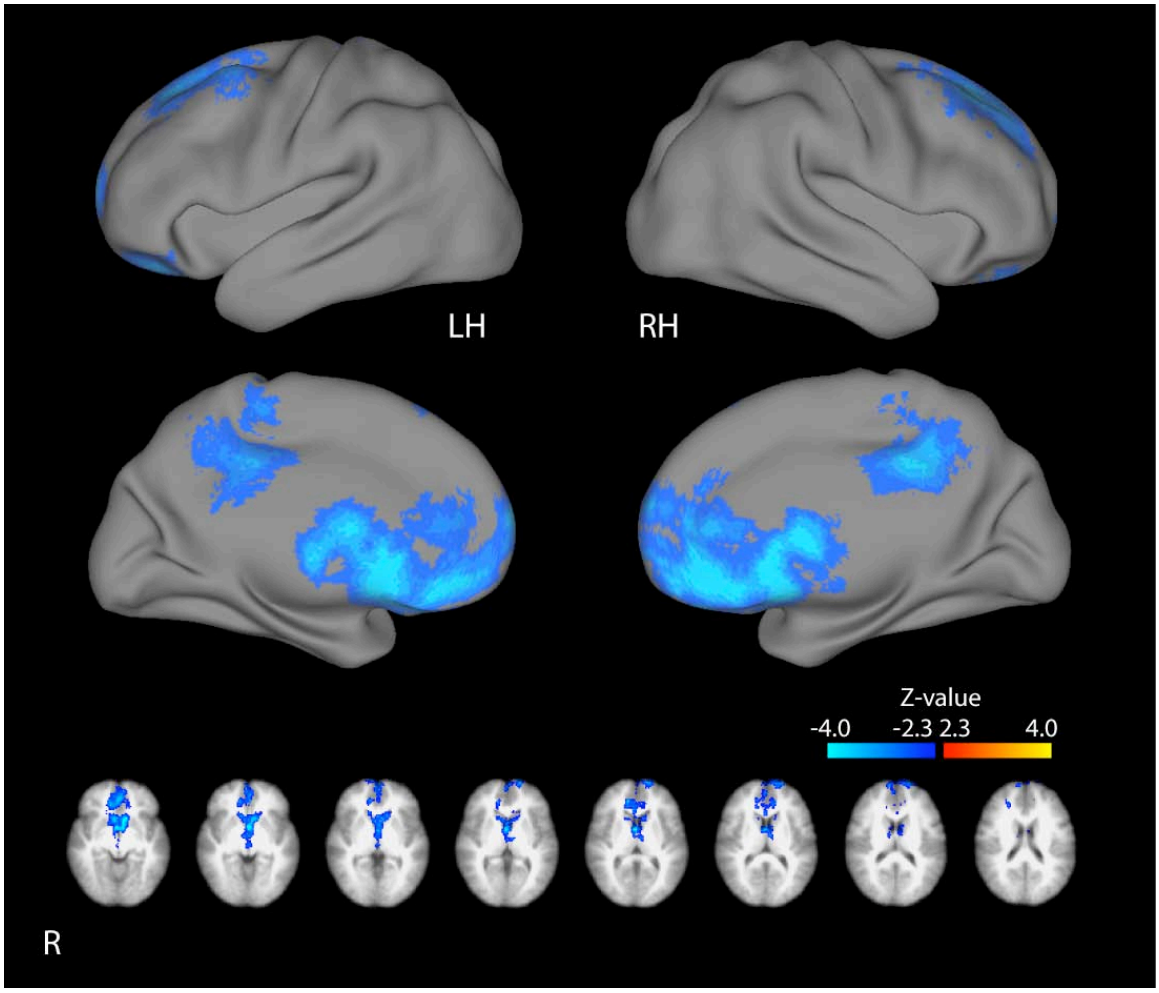
Supplementary Figure 1. Regions with significant activation for task vs. baseline (Z > 2.3, whole-brain cluster-corrected at p < .05 using GRFT). Red-yellow scale reflects positive activation, blue-white scale reflects negative activation.



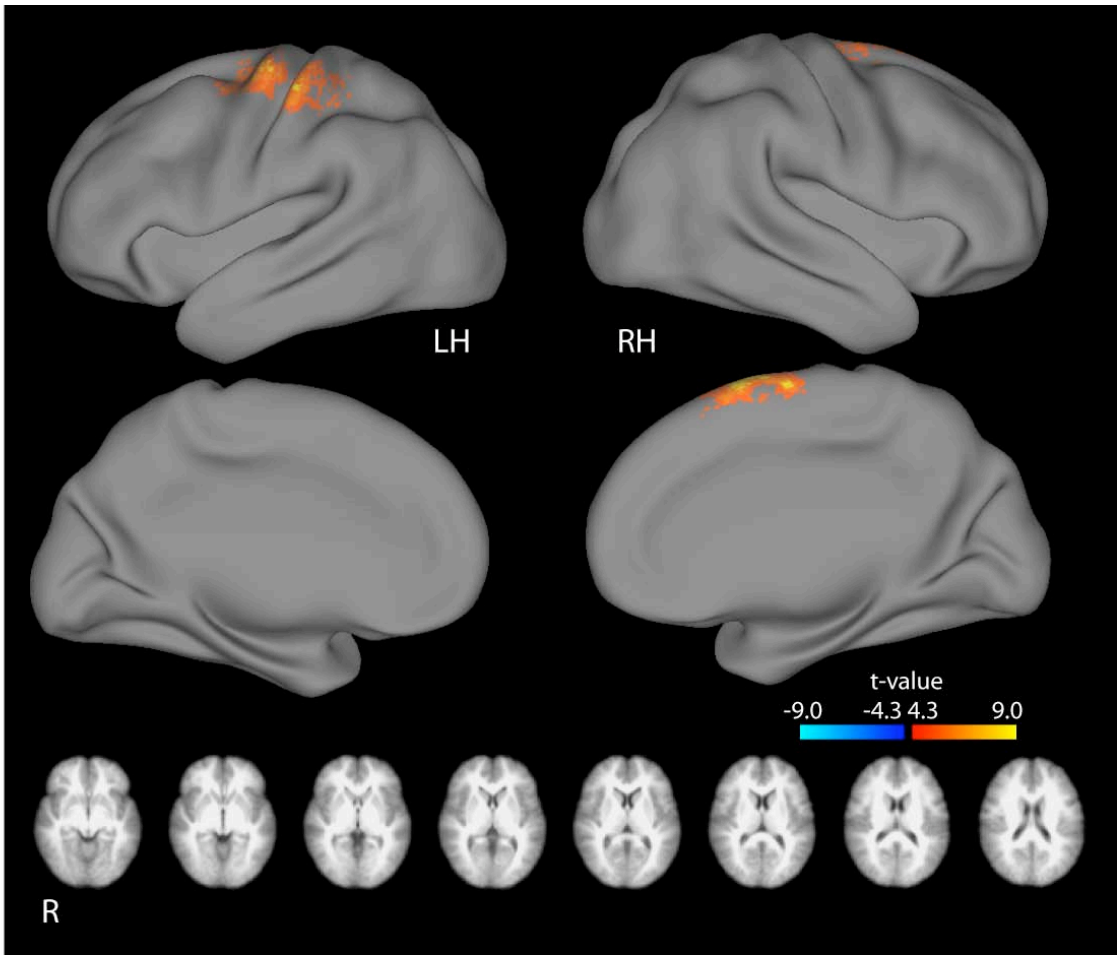
Supplementary Figure 2. Regions with significant parametric increase in fMRI signal to increasing potential gains (Z > 2.3, whole-brain cluster-corrected at p < .05 using GRFT). No regions showed decreasing activity for increasing potential gains.



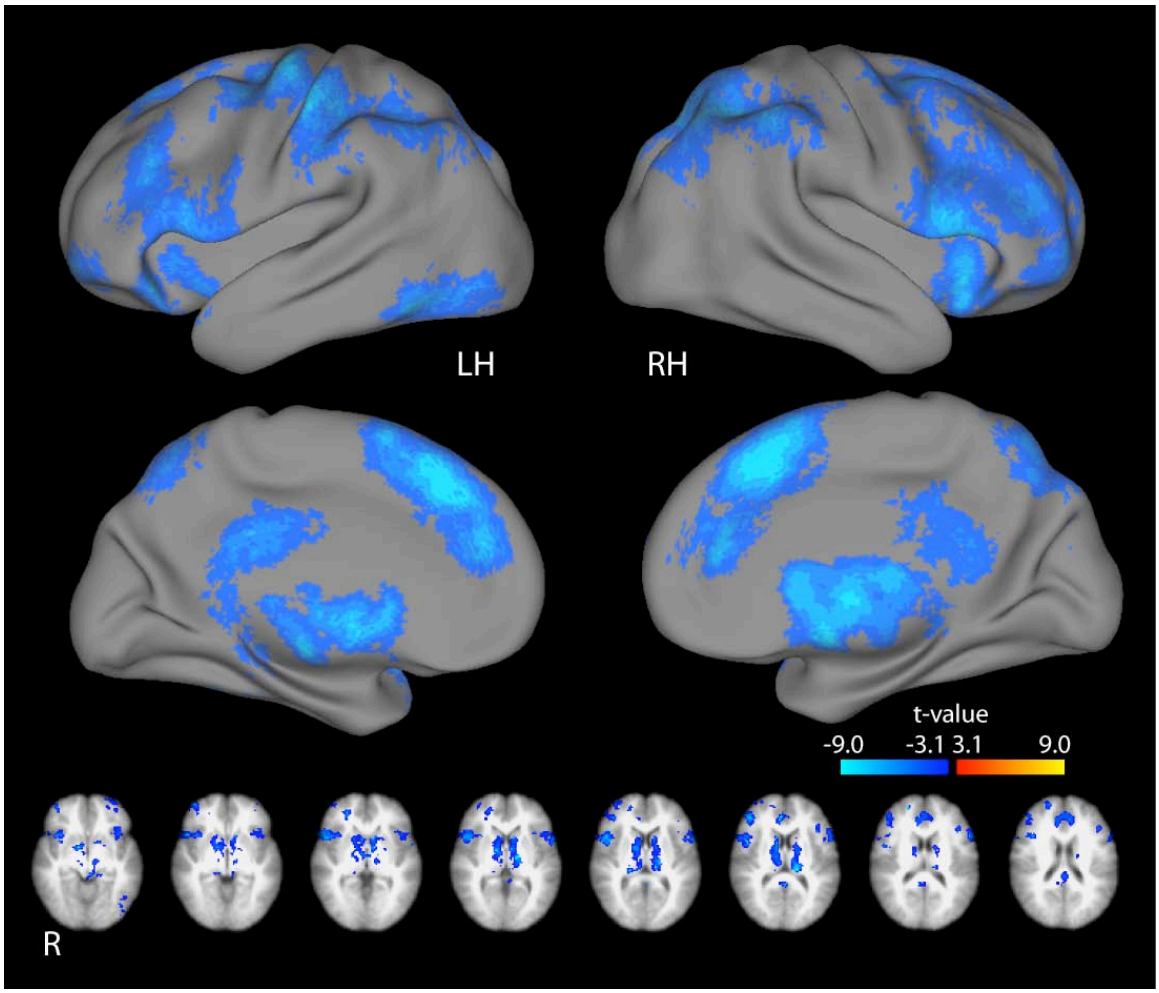
Supplementary Figure 3. Regions with significant parametric decrease to increasing potential losses (Z > 2.3, whole-brain cluster-corrected at p < .05 using GRFT). No regions showed increasing activity for increasing potential losses.



Supplementary Figure 4. 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.



Supplementary Figure 5. 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.



Supplementary Figure 6. 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.