NeuroImage 64 (2013) 582–589

Contents lists available at SciVerse ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Counterfactual thinking and reward processing: An fMRI study of responses to gamble outcomes

Sarah E. Henderson *, Catherine J. Norris

Dartmouth College, USA

ARTICLE INFO

Article history:
Accepted 28 August 2012
Available online 5 September 2012

Keywords:
Counterfactual thinking
Reward
Uncertainty
Gambling
fMRI
Emotion

ABSTRACT

The process of comparing obtained outcomes to alternative possible outcomes, known as counterfactual thinking, is inescapable in daily life; however, the neurocognitive mechanisms underlying counterfactual thinking and how they influence emotional responses to better and worse outcomes is not well understood. We conducted an event-related functional magnetic resonance imaging (fMRI) gambling study in which participants were informed of two equally possible outcomes of a card gamble before they selected a card. Participants reported experiencing mixed emotions (i.e., both positive and negative affect) for disappointing wins (winning the lesser of two amounts) and relieving losses (losing the lesser of two amounts). Neuroimaging results supported the hypothesis that these mixed emotions were associated with activation of a fronto-parietal network, which subsequently influenced processing in reward and punishment regions (dorsal and ventral striatum, right anterior insula). The fronto-parietal network was sensitive to outcomes that resulted in mixed emotions, whereas reward and punishment regions were sensitive to comparisons between obtained and unobtained outcomes. These findings provide insight into the neurocognitive mechanisms underlying the mixed emotional experiences that result from counterfactual comparisons, and inform our understanding of how the brain is optimized to use the wealth of environmental information to inform current and future behavior.

Introduction

Imagine you are anxiously awaiting news about your yearly bonus when you learn that a colleague in your pay grade has received a $1000 bonus. Your check arrives in the sum of $500: how do you feel? Most people would likely feel disappointment despite gaining money, given the knowledge of a better possible outcome and the anticipation of the receipt of this better outcome. This example illustrates the process of counterfactual thinking, in which emotions are influenced by comparisons of an obtained outcome to an unobtained outcome (Kahneman and Miller, 1986). In this way, emotional responses to the exact same outcome (e.g., obtaining $0) can vary greatly depending on unobtained outcomes (winning $10 or losing $10; Mellers et al., 1997). Researchers have used a variety of mixed gamble outcomes and consistently find that participants report disappointment or regret when obtained outcomes are worse than unobtained outcomes, and elation or relief when obtained outcomes are better than unobtained outcomes (Chua et al., 2009; Larsen et al., 2004; Mellers and McGraw, 2001; Mellers et al., 1999). These experiences of disappointment and relief are attributed to the occurrence of mixed emotions in which one simultaneously experiences positive and negative affect (i.e., ambivalence; Larsen et al., 2001). Mixed emotions, therefore, are likely to emerge in situations in which an unobtained comparison outcome influences emotional responses to an obtained outcome. Little is known about how counterfactual comparisons are processed in the brain, and how such patterns of brain activation impact subsequent emotions.

An understanding of the neural processes underlying counterfactual thinking could help explain how and why our emotional responses are impacted by unobtained outcomes. Ursu and Carter (2005), focusing on the role of the orbitofrontal cortex (OFC) in reward processing, found greater medial OFC (BA 10) activity when participants received outcomes that were comparatively more rewarding, and greater lateral OFC (BA 11/47) activity for outcomes that were comparatively more penalizing. The ventral and dorsal striatum have also been shown to be implicated in comparisons between obtained and unobtained outcomes, with greater activity for comparatively better outcomes that produce relief, and deactivations for comparatively worse outcomes that produce regret (Breiter et al., 2001; Nicolle et al., 2011). Interestingly, in the context of more positive comparisons, win outcomes elicited less activation of the dorsal striatum than did loss outcomes, indicating that there may be an asymmetry in processing wins and losses in these regions when they involve counterfactual comparisons.

* Corresponding author at: Department of Psychological and Brain Sciences, Dartmouth College, 6207 Moore Hall, Hanover, NH 03755, USA.
E-mail address: sarah.e.henderson@dartmouth.edu (S.E. Henderson).

1053-8119/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neuroimage.2012.08.078

Other researchers have focused on the neural correlates of the consequences of counterfactual thinking, mainly the experience of regret and disappointment. Coricelli et al. (2005) found that the experience of regret when an obtained monetary outcome was less than the best possible outcome was associated with increased activity in the medial OFC, lateral OFC, dorsolateral PFC, and inferior parietal lobule (IPL). The authors interpreted their results as indicating that counterfactual thinking activates a network of regions involved with evaluations of complicated emotions (medial OFC), reversal learning (lateral OFC), online monitoring of decision-making (dorsolateral PFC), and action desirability (IPL). In another relevant neuroimaging study, Chua et al. (2009) found that activation of the lateral OFC, dorsolateral prefrontal cortex (DLPFC), and sublenticular extended amygdala/ ventral striatum increased as the difference between obtained and unobtained monetary outcomes increased. Additional analyses showed that both disappointment and regret led to increased activity in the anterior insula and dorsomedial PFC, whereas regret increased activity in the right lateral OFC and superior frontal gyrus (BA 8; Chua et al., 2009). These findings suggest that, in addition to striatal and orbitofrontal regions, dorsal prefrontal and parietal regions may be involved in processing counterfactual comparisons and the emotions that they generate.

Overall, the neuroimaging literature on counterfactual thinking and mixed emotions implicates regions involved in the processing of affective salience and reward/punishment, as well as regions of prefrontal cortex implicated in decision-making and performance evaluation, but falls short of providing an understanding of how activity in different neural networks might influence each other and ultimately relate to experienced emotions. The current study sought to identify neural networks involved in counterfactual comparisons and how they relate to the emotional experience of these outcomes. In order to do this, we used a two-card gambling paradigm (modified from Larsen et al., 2004, 2009) in which participants select a card with full knowledge of the two possible outcomes, and then learn which outcome was obtained versus unobtained. We first conducted an independently-selected ROI analysis to investigate if regions previously found to be responsive to financial gains (e.g., dorsal striatum, MOFC) and losses (e.g., anterior insula) would be more sensitive to the outcome (i.e., simply winning or losing money, regardless of the comparison); or would be more sensitive to the comparison (i.e., obtaining better or worse outcomes). We predicted that these regions would be more sensitive to the comparison than the outcome given previous research demonstrating that these regions (e.g., dorsal striatum) respond differently to the same value when comparison values differed (Breiter et al., 2001; Nicolle et al., 2011). Next, we conducted a whole-brain analysis of variance (ANOVA) and expected to find increased activation for gamble outcomes that produce mixed emotions in a network of prefrontal (e.g., DMPFC, DLPFC), subcortical, and parietal regions previously implicated in monetary decision-making tasks (Chua et al., 2009; Coricelli et al., 2005). We then examined relationships between functional brain activation and emotional responses to counterfactual outcomes and expected that increased activation in the observed prefrontal, subcortical, and parietal regions would correlate with greater mixed emotions. Ultimately, we aimed to provide a more complete explanation of the neural mechanisms underlying counterfactual thinking and the experience of mixed emotions that result from this comparison process.

Methods

Participants

Thirty (15 female) right-handed, native English-speaking Dartmouth students between the ages of 18 and 25 ($M = 19.47$, $SD = 1.63$) were recruited through a Dartmouth listserve and received either extra credit for an introductory psychology class or monetary compensation for their participation. All participants provided informed consent and the study was conducted under the guidance of the Dartmouth College Committee for the Protection of Human Subjects. All participants were screened for medications that could potentially impact emotional processing (e.g., antidepressants), a history of head injuries, and other psychological (e.g., bipolar disorder) and physical conditions (e.g., pacemaker) that impact typical emotional processing or would be unsafe in the MRI environment.

Gambling task

After participants gave informed consent, they completed an MRI safety questionnaire and answered a few questions about their prior gambling experiences. Participants were informed at this point that they would receive an endowment of $10 to gamble with during the current study. They were told that they could increase or decrease the endowment based on their gambling choices. Gambling outcomes, however, were predetermined to ensure that all participants experienced the same outcomes over the course of the study. Thus, all participants won $10 and received this amount at the end of the session. During debriefing we probed for suspicion that the outcomes had been predetermined. Six participants reported some suspicion that the gamble outcomes were predetermined; however, all analyses were conducted with and without these participants and there were no differences for any of the findings. Therefore, data from all 30 participants are reported.

While in the MRI scanner, participants played a series of gambles in which they chose between two cards which were associated with monetary gains or losses (Larsen et al., 2004; see Fig. 1 for a trial schema). Trials began with a fixation cross presented for 0.5 s and were followed by the appearance of two facedown cards and the monetary gains or losses associated with the pair of cards, which were presented for 2 s. In each scenario, both cards were either win cards (e.g., win $18 vs. win $36) or loss cards (e.g., lose $18 vs. lose $36). Following presentation of the stakes, participants were given 2 to 12 s to select either the left or right facedown card using a two-button box. The choice period was jittered to allow for deconvolution of the neural responses to the onset of the possible outcomes. The selected card was outlined in red. Finally, the cards flipped over and participants viewed the outcome of their choice for 4 s, followed by a variable inter-trial interval ranging from 1.5 to 13.5 s.

There were four possible outcomes for each gamble: (1) “Outright Win,” winning the larger amount of two outcomes (e.g., $36 instead of $18), (2) “Disappointing Win,” winning the smaller amount of two outcomes ($18 instead of $36), (3) “Relieving Loss,” losing the smaller amount of two outcomes, and (4) “Outright Loss,” losing the larger amount of two outcomes. Overall, there were 32 unique card outcomes in the study (8 for each outcome type), plus 4 unique filler outcomes, for a total of 36 possible outcomes. The experiment consisted of four blocks of 27 trials each comprised of 12 wins, 12 losses, and 3 filler trials. The card choice scenarios were presented in one of two predetermined random orders which were counterbalanced across participants.

An HP laptop running E-prime 2.0 Professional presented gambles and recorded participants' choices and response times in the scanner. Visual images were reflected through mirrors mounted on the head coil and projected onto a screen located behind the scanner. Button press responses for card choices were made on an fMRI-compatible response box.

Post-task gamble ratings

Following the scanner portion of the study, participants were seated at a computer to rate their emotional responses to each

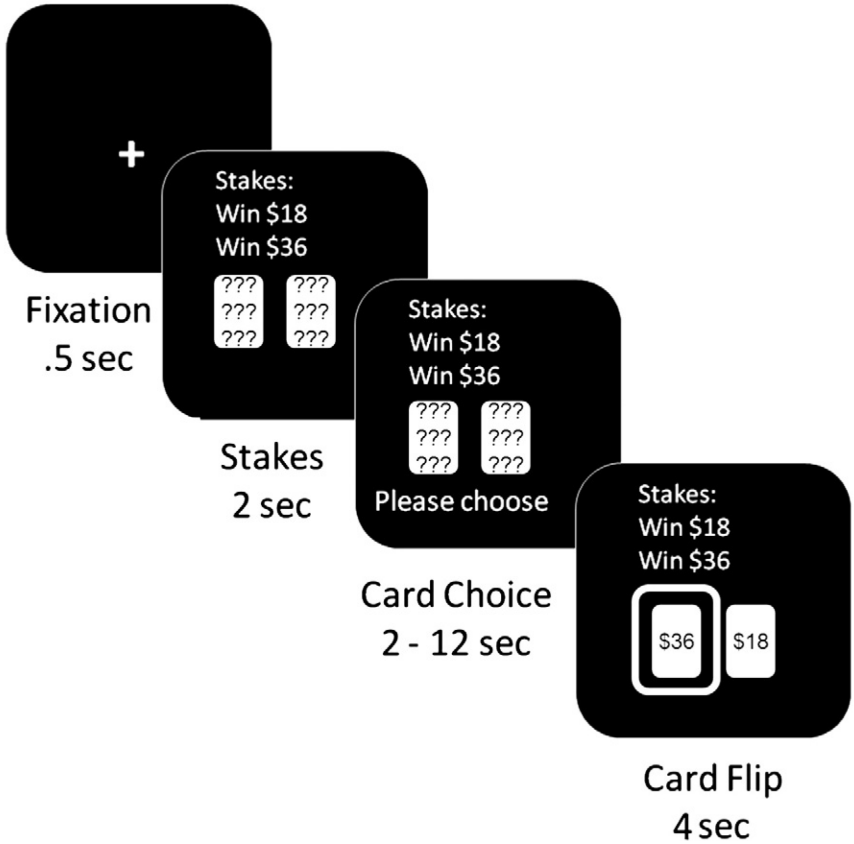

Fig. 1. Trial progression in which participants viewed a fixation, followed by presentation of the trial stakes, were then asked to select a card, and eventually viewed the outcome of their selection. Note that in the scanner the selected card was outlined in red.

gamble outcome.[1] Participants viewed a fixation cross for 0.5 s followed by one of the 36 unique gamble outcomes that they received in the scanner, which remained visible until they made a rating using the Evaluative Space Grid (ESG; Larsen et al., 2009). The ESG is a 5×5 grid that allows for simultaneous ratings of positivity on the *x*-axis (0 = *not at all positive*; 4 = *extremely positive*) and negativity on the *y*-axis (0 = *not at all negative*; 4 = *extremely negative*). For each gamble outcome, participants moved the mouse to the cell that most accurately captured their emotional responses. After each rating, there was a 2 s inter-trial interval preceding the next trial.

fMRI data acquisition

Images were acquired on a 3-T Philips Achieva Intera scanner with an eight-channel SENSE head coil. High-resolution T1-weighted anatomical images were acquired in 160 1-mm sagittal slices (TR = 9.8 ms (shortest), TE = 4.6 ms, flip angle = 8°, field of view = 24 cm). Functional T2*-weighted gradient echo images were acquired in four runs with 21 contiguous 5.0 mm axial interleaved slices with a 1.0 mm gap (TR = 2000 ms, TE = 35 ms, flip angle = 90°, FOV = 24 cm, 80×80 matrix size, fat suppressed, level 1 spoiler gradient).

fMRI data processing and analyses

fMRI data were preprocessed and analyzed using the AFNI software package (Cox, 1996). We first despiked the data and corrected for slice-timing acquisition using an in-house correction algorithm. Images were registered to a volume collected in the middle of the scanning session, and two general linear models (GLM) were performed: the first removed motion-related variance, and the second used the residual variance to examine condition-related differences. In the first GLM, we regressed out the six motion parameters. We then examined the standard deviations of baseline estimates to ensure adequate signal to noise ratios. Data were smoothed with a 6-mm FWHM Gaussian kernel. The second GLM included our four gamble outcomes as regressors of interest, as well as one regressor of no interest to model the anticipation period for each trial, and four regressors of no interest to remove variance associated with linear trends and signal drift. Specifically, this GLM estimated task-related changes in blood oxygen level dependent (BOLD) signal at each voxel by modeling the onsets for outright wins, disappointing wins, relieving losses, and outright losses. All responses were modeled as sine waves with three base functions, convolved with a hemodynamic response to form individual subject contrast maps. Finally, these subject maps were co-registered to in-plane anatomical images, resampled to $1\times1\times1$ mm voxels, and normalized to standard Talairach space using the ICBM 452 template.

We first conducted a series of region of interest (ROI) analyses on independent ROIs that had previously been found to be sensitive to monetary wins and losses (Liu et al., 2007; wins: right dorsal striatum (18, 16, 4), left dorsal striatum (−20, 8, −4), left medial OFC (−20, 34, −16), right medial OFC (20, 34, −16), right inferior frontal gyrus (34, 30, 26), left inferior frontal gyrus (−34, 30, 26), and left superior frontal (−24, 32, 36; −16, 6, 52); losses: left anterior insula (−40, 16, −14), right anterior insula (42, 14, −18), left inferior frontal gyrus (−60, 20, 20), right inferior frontal gyrus (52, 34, 12), right DMPFC (2, 54, 30), left DMPFC (−2, 54, 30), and midbrain (0, −30, 2). For each ROI, a 5 mm sphere was placed at the reported coordinates (at the center of mass for each cluster) and a bilateral region for the two medial clusters to ensure we had adequate coverage. Parameter

[1] Ratings were collected post-scanning due to considerations about increasing the amount of time participants spent in the scanner, and also due to the variable amount of time it takes individuals to use the ESG.

estimates were extracted and submitted to offline analyses in SPSS. ROI analyses were False Discovery Rate (FDR) corrected for multiple comparisons, using a q-value of 0.1 (Benjamini and Hochberg, 1995).

Whole-brain ANOVAs were subsequently conducted in AFNI and parameter estimates (β) for significant clusters were extracted and submitted to offline analyses in SPSS. All analyses were conducted using area under the curve (AUC; 5 to 9 s) indices. To correct for multiple comparisons, we ran a Monte Carlo simulation on an averaged group mask with 1000 iterations, assuming 6 mm of interdependence (based on our smoothing parameter). Using an individual voxelwise threshold of $p=.01$, a minimum cluster volume of 970 μL resulted in a whole brain correction to $p<.05$. Finally, a series of two-tailed Pearson correlations were performed between behavioral ratings of affect and functional imaging data from the whole-brain analysis and ROI analysis.

Results

Affect ratings and response times

In order to examine emotional responses to the gamble outcomes, positive and negative ratings from the ESG were subjected to separate 2(outcome: win, loss)∗2(comparison: better, worse) ANOVAs. The ANOVA conducted on positive ratings revealed both the expected main effect of outcome, $F(1,29)=150.35$, $p<.001$, such that win outcomes were rated more positively ($M=2.01$, $SD=0.52$) than loss outcomes ($M=0.61$, $SD=0.40$), as well as the main effect of comparison, $F(1,29)=74.64$, $p<.001$, such that better outcomes were rated more positively ($M=1.71$, $SD=0.51$) than worse outcomes ($M=0.91$, $SD=0.32$; Fig. 2). The outcome × comparison interaction was also significant, $F(1,29)=14.30$, $p<.005$. Pairwise comparisons revealed that participants reported the most positive affect for outright wins ($M=2.30$, $SD=0.58$), followed by disappointing wins ($M=1.71$, $SD=0.59$), relieving losses ($M=1.12$, $SD=0.70$), and outright losses ($M=0.10$, $SD=0.21$), all pairwise $ps<.001$.

Similarly, the ANOVA for negative ratings revealed a main effect of outcome, $F(1,29)=130.26$, $p<.001$, such that losses were rated more negatively ($M=2.03$, $SD=0.46$) than wins ($M=0.60$, $SD=0.45$), as well as a main effect of comparison, $F(1,29)=84.00$, $p<.001$, such that worse comparisons were rated more negatively ($M=1.72$, $SD=0.45$) than better comparisons ($M=0.91$, $SD=0.30$; Fig. 2). The expected interaction was also significant, $F(1,29)=9.63$, $p<.005$, with pairwise tests revealing that negative affect was highest for outright losses ($M=2.34$, $SD=0.48$), followed by relieving losses ($M=1.73$, $SD=0.54$), disappointing wins ($M=1.10$, $SD=0.76$), and outright wins ($M=0.10$, $SD=0.27$), all pairwise $ps<.001$. Taken together, the patterns of findings for both positive and negative ratings suggest that although winning or losing clearly impacted emotional responses, the comparison of the obtained outcome to the unobtained outcome also had an influence on emotional responses.

In order to determine if participants were truly experiencing both positive and negative affect simultaneously, we calculated participants' minimum (MIN) ratings for each outcome by taking the lower value of the positive and negative ratings for a single gamble (Schimmack, 2001). In this way, a high MIN score indicates greater mixed emotions (e.g., if positivity = 4 and negativity = 3, MIN = 3) and a low MIN scores indicates fewer or no mixed emotions (e.g., if positivity = 4 and negativity = 0, MIN = 0). We submitted MIN ratings to a 2(outcome: win, loss) × 2(comparison: better, worse) ANOVA and found no main effects of outcome or comparison, but a significant interaction, $F(1,29)=69.15$, $p<.005$ (Fig. 2). Pairwise tests revealed that disappointing wins ($M=0.85$, $SD=0.55$) and relieving losses ($M=0.86$, $SD=0.56$) elicited equal amounts of mixed emotions, and outright wins ($M=0.10$, $SD=0.24$) and outright losses ($M=0.10$, $SD=0.18$) elicited equal amounts of mixed emotions (i.e., very little). Importantly, both disappointing wins and relieving losses had higher MIN scores than outright wins or outright losses, $ps<.001$.

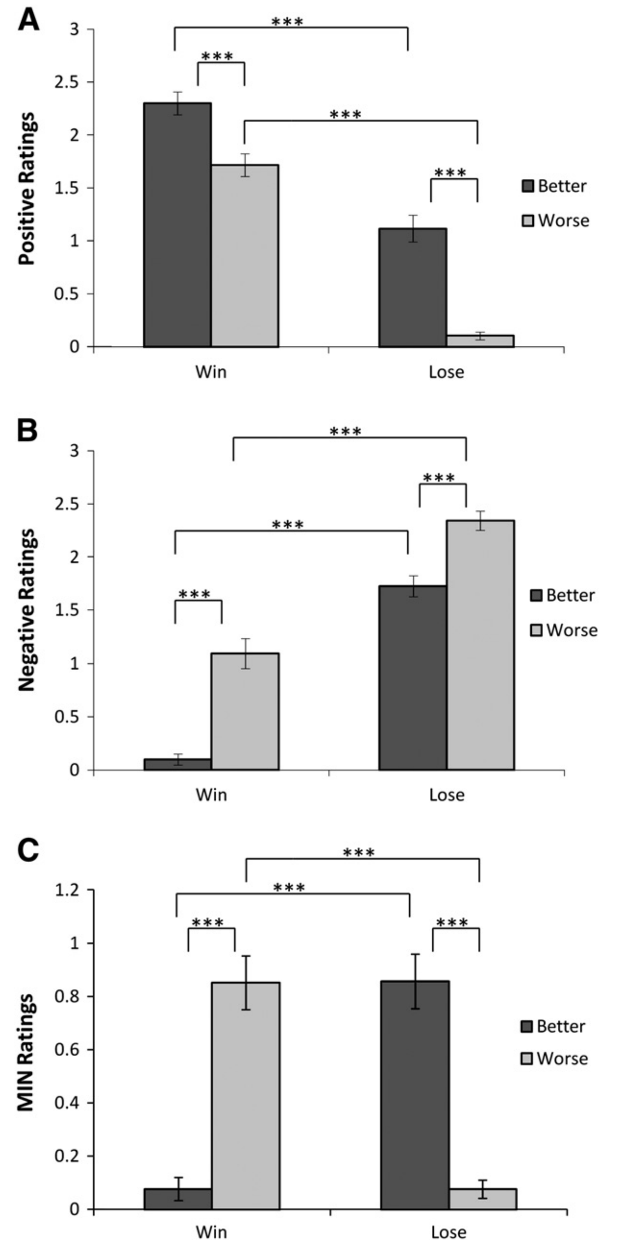

Fig. 2. Behavioral ratings for each of the four trial outcomes for (A) positivity, (B) negativity, and (C) minimum ratings (i.e., ambivalence or mixed emotions).

In order to investigate if card choice times differed as a function of the valence of the trial outcome (i.e., win or loss trial), we first trimmed the top and bottom 5% of responses (Bush et al., 1993) and then submitted these to a paired-samples t-test. Participants were significantly faster when choosing a card in the anticipation of a win ($M=815.81$ ms, $SD=263.21$ ms) than the anticipation of a loss ($M=885.37$ ms, $SD=261.39$ ms), $t(29)=2.58$, $p<.05$.

Analyses of reward and punishment ROIs

In order to investigate a priori hypotheses regarding the modulation of neural responses to obtained versus unobtained monetary outcomes, we submitted estimates of neural activation from each ROI to 2(outcome: win, loss)∗2(comparison: better, worse) ANOVAs. We expected to find activity in these regions that was sensitive to

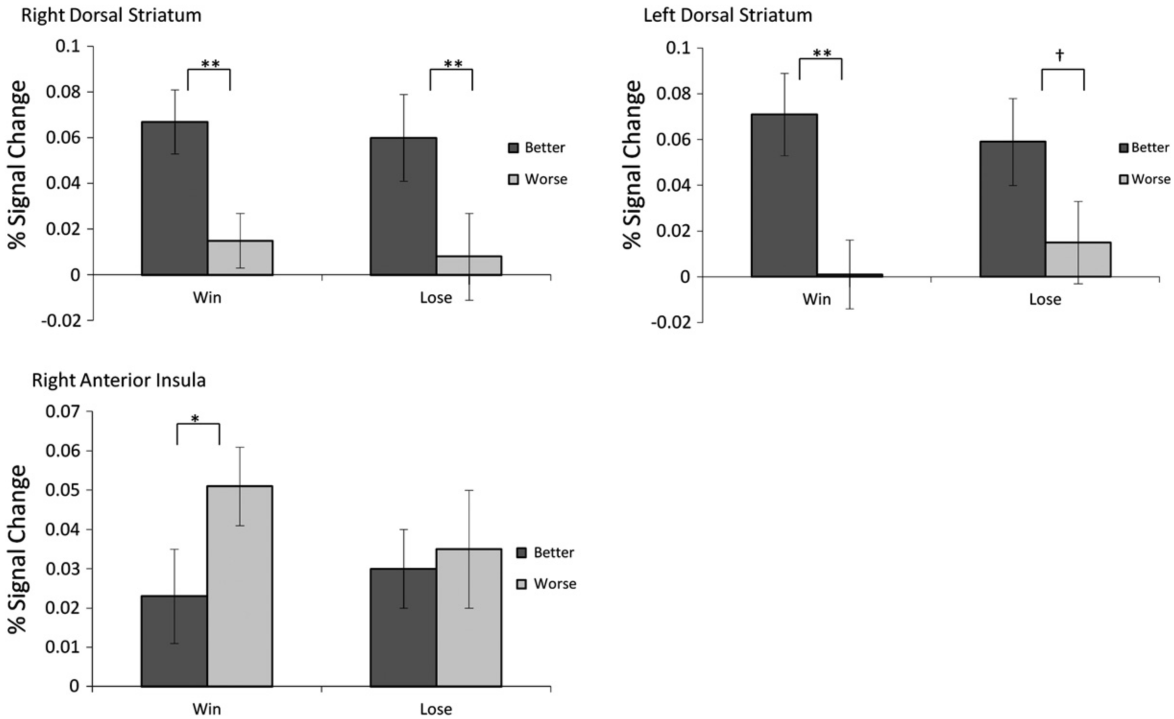

Fig. 3. Percent signal change from ROI analysis demonstrating significant main effects for comparison in the right dorsal striatum, left dorsal striatum, and right anterior insula.

comparisons between obtained and unobtained outcomes rather than sensitive to the outcome of winning or losing money overall. We found a main effect of comparison in the right and left dorsal striatum, such that these regions were more activated by outcomes that were better than those that were worse by comparison, $F(1, 29)=22.56$, $p<.001$; $F(1, 29)=14.56$, $p<.005$, respectively (Fig. 3). Pairwise tests confirmed that activity was greater for outright wins than disappointing wins, and greater for relieving losses than outright losses (all pairwise $ps<.005$, except $p<.06$ for the left dorsal striatum when comparing relieving losses to outright losses).

For ROIs that have previously been shown to respond to losses, a region in the right anterior insula showed a main effect of comparison, with greater activation to worse than better comparisons, $F(1, 29)=5.29$, $p<.05$ (Fig. 3). Pairwise tests showed that activity for disappointing wins was significantly greater than outright wins ($p<.05$), whereas there was no difference between relieving losses and outright losses.[2]

Neural responses to mixed versus outright monetary outcomes

We also conducted a whole brain 2(outcome: win, loss)∗2(comparison: better, worse) ANOVA on estimates of neural activation to examine the involvement of other regions in counterfactual thinking. Although no clusters survived correction for multiple comparisons for the main effect of outcome, the main effect of comparison revealed two clusters in the right and left NAcc, whole brain corrected to $p<.05$ (Table 1, Fig. 3). Beta weights revealed that outcomes that were better by comparison (i.e., outright wins, relieving losses) were associated with greater activation in the bilateral NAcc than did outcomes that were worse by comparison (i.e., disappointing wins, outright losses). Examination of pairwise comparisons for both clusters found that activity for outright wins was significantly greater than disappointing wins, and activity for relieving losses was significantly greater than outright losses (pairwise $ps<.005$).

The whole brain analysis also revealed an outcome × comparison interaction in several clusters in parietal and frontal regions, whole brain corrected to $p<.05$ (Table 1). Three of these clusters, located in the left and right posterior parietal cortex (PPC; BA 40) and in dorsolateral prefrontal cortex (DLPFC; BA 6/8), exhibited similar patterns of activation. Each of these regions showed greater activation for mixed outcomes (i.e., disappointing wins, relieving losses) than for univalent outcomes (i.e., outright wins, outright losses; Fig. 4). Specifically, the right and left PPC were more active for disappointing wins than outright wins and outright losses, and were more active for relieving losses than outright wins (but not outright losses). The right DLPFC cluster showed greater activity for disappointing wins than outright wins and outright losses. In sum, these regions were more active for outcomes that produced mixed emotions than those that produced pure positivity or negativity, with disappointing wins producing the most robust effects, and greater activity for relieving losses than outright wins and outright losses.

Table 1
Peak activations in outcome × comparison ANOVA.

Region	BA	F-value	X	Y	Z
Comparison: better > worse					
L nucleus accumbens		12.36	−15	−14	−2
R nucleus accumbens		12.23	13	−15	−1
Outcome × comparison interaction					
R posterior/inferior parietal sulcus	40	14.78	35	−59	48
L posterior parietal	40	9.70	−39	51	50
R premotor to dorsomedial PFC	6/8	8.30	32	−19	53
R superior temporal gyrus		8.32	57	35	14
R insula		8.93	38	6	9

Note: clusters of 970 or more contiguous voxels whose global maxima meets an F threshold of 7.10 ($p<.01$, corrected to $p<.05$) are reported. The center of mass x, y, z coordinates are in Talairach coordinates. BA = Brodmann's area. L = left; R = right.

[2] It is worth noting that other regions from the ROI analysis showed significant outcome∗interaction effects, the details of which go beyond the scope of this paper, but no regions showed a main effect of outcome.

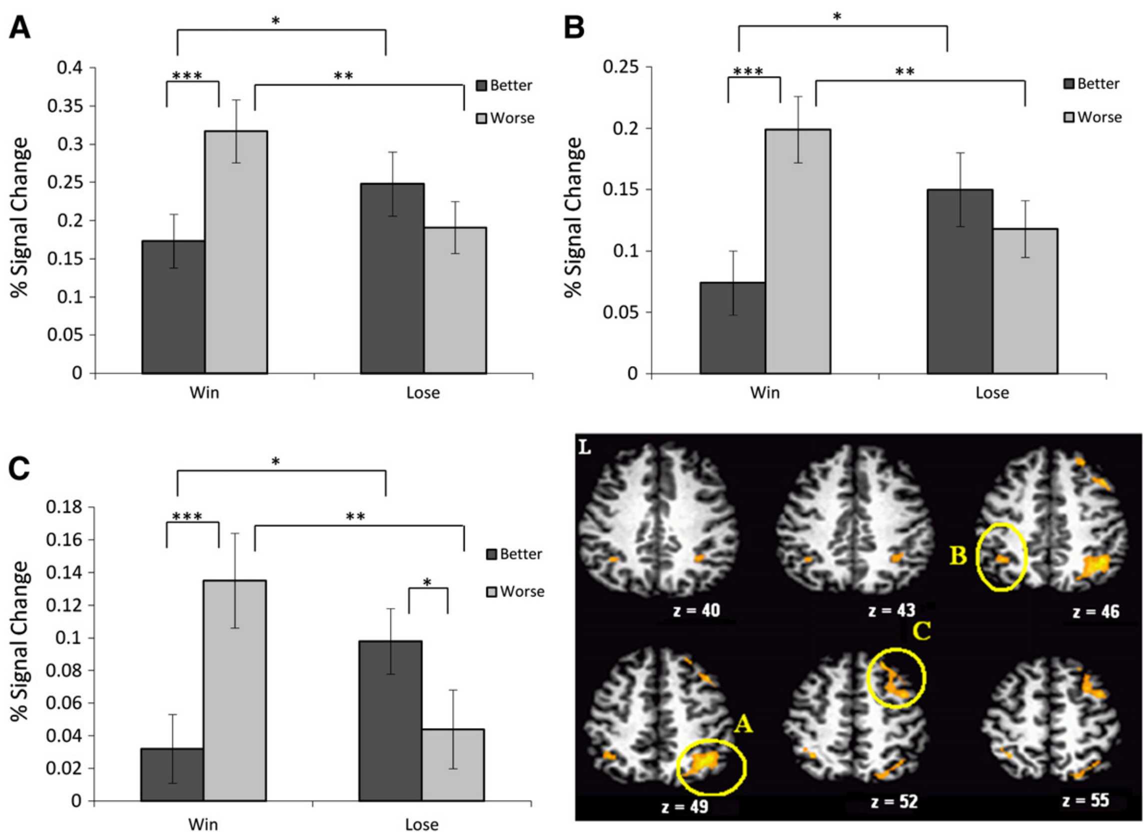

Fig. 4. Percent signal change for outcome × comparison interaction regions for each of the four trial outcomes for (A) right posterior parietal cortex, (B) left posterior parietal cortex, and (C) right dorsolateral prefrontal cortex.

Correlations of neural activation with affect ratings

A final aim of the current study was to conduct an exploratory analysis of the relationships between patterns of neural activity and the mixed emotions experienced in response to relieving losses and disappointing wins. We correlated MIN ratings for relieving losses and disappointing wins with estimates of neural activity for these gamble outcomes in the three clusters that emerged from the outcome × comparison interaction (i.e., bilateral PPC, right DLPFC). Surprisingly, neural activity in these regions did not correlate with self-reported mixed emotions for either relieving losses or disappointing wins; however, it is possible that these neural regions exert their influence on emotions indirectly through connections with other networks, particularly those involved in reward and punishment. To examine this possibility, we correlated mean neural activity in regions implicated in counterfactual thinking (i.e., bilateral PPC, right DLPFC) with comparison-related activity from our whole-brain analysis and a priori ROIs (i.e., bilateral NAcc, bilateral dorsal striatum, right anterior insula). When participants experienced relieving losses, activity in the right PPC was positively correlated with activity in the left dorsal striatum, right NAcc and left NAcc, $rs(28) = 0.49$, 0.54 and 0.44, $ps < .05$, respectively.[3] Similarly, activity in the right DLPFC was positively correlated with activity in the right anterior insula, $r(28) = 0.38$, $p < .05$. Overall, these results suggest that activation of neural regions associated with counterfactual thinking did indeed correlate with activation of subcortical and prefrontal structures that process reward and punishment information.

Given these findings, we lastly sought to examine whether activity in reward regions correlated with participants' emotional responses. For relieving losses, neural activation of reward regions was positively correlated with MIN ratings for the left dorsal striatum, left NAcc, and right anterior insula, $r(28) = 0.45$, 0.45 and 0.40, $ps < .05$ respectively. Importantly, MIN ratings are driven by non-dominant responses to outcomes (i.e., positivity ratings on a loss outcome are non-dominant, whereas negativity ratings are dominant), which suggests that activation of reward regions is associated with increased positivity in response to losses, which is brought about through counterfactual comparisons. The same correlations were conducted for disappointing win outcomes. The pattern of correlations between regions sensitive to counterfactual thinking and reward regions was the same as for relieving losses, although only one correlation reached traditional levels of significance. Activity in the right NAcc, was positively correlated with MIN ratings for disappointing wins, $r(28) = 0.44$, $p < .05$, again suggesting a relationship between activity in subcortical regions and the emotional experience of counterfactual comparisons.

Discussion

The primary aim of the current study was to examine the neural mechanisms underlying mixed emotional responses experienced as a function of counterfactual thinking, with the hope that this investigation would ultimately shed light on our understanding of the psychological processes engaged in counterfactual comparisons. We found regions of the parietal and frontal cortices that were responsive to mixed outcomes, which positively correlated with activation in regions implicated in reward and punishment. These reward and punishment regions were sensitive to better and worse comparisons, and activity in these regions positively correlated with reported feelings of emotional conflict for mixed outcomes.

[3] It is worth noting that the left PPC showed the same correlation patterns as the right PPC but each failed to reach significance.

Outcome processing and counterfactual comparisons

Our results suggest that the emotions brought about by counterfactual comparisons are carried out by the fronto-parietal network, which in turn influences activation in reward and punishment processing regions to allow for consideration of comparison values before making the ultimate affective evaluation. As the fronto-parietal network has been implicated in numerical comparisons and calculations (Hirsch et al., 2001; Pinel et al., 2001), attention and action preparation (Behrmann et al., 2004; Creem-Regehr, 2009), and working memory (Mars and Grol, 2007), the observed pattern of activity may reflect that conflicting outcomes either bring online some form of higher-level cognitive process to carry out value comparisons or that conflicting outcomes require greater neural resources to make appropriate responses. We also found that better comparisons drove increases in activity in regions of the striatum; whereas worse comparisons drove increases in activity in the anterior insula. Thus, our results also provide support that positive and negative information are processed in distinct, separable neural systems. Positive counterfactual comparisons impacted circuits involved with reward learning (Breiter et al., 2001; Chua et al., 2009; Delgado, 2007; Nicolle et al., 2011), whereas negative counterfactual comparisons activated a region implicated in emotional experience and expression (Aggleton, 1992; Augustine, 1996; Fink et al., 1996) and negative monetary outcomes (Chua et al., 2009; Liu et al., 2007). The current findings are consistent with research showing that frontal, parietal, and subcortical regions work together to influence emotional responses to counterfactual comparisons. This comes from diffusion tensor imaging research, which has shown strong connectivity between these regions (Uddin et al., 2010).

Differences for wins and losses

Although the observed neural and behavioral patterns for wins and losses were very similar, we did find a few differences worth further discussion. First, we found that activity in reward and punishment regions distinguished between outright wins and disappointing wins, but not as robustly between outright losses and relieving losses. One possible explanation comes from the contagion theory of the negativity bias (Rozin and Royzman, 2001) which posits that the effect of a negative addition to an overall positive situation (e.g., adding a spoonful of tar to a vat of honey) is far more salient than the effect of a positive addition to a negative situation (e.g., adding a spoonful of honey to barrel of tar). In other words, the experience of negativity (i.e., disappointment) in the context of a win has a stronger effect on neural responses than does the experience of positivity (relief) in the context of a loss.

Another difference between wins and losses was that correlations between neural activation and emotion ratings were more robust for relieving losses than for disappointing wins. One possible explanation could be that a loss outcome may elicit more self-regulatory processing, that participants may attempt to find and focus on any positive aspect of a negative outcome to alleviate that felt negativity (Kassam et al., 2011). Thus, relieving losses may produce comparably stronger relationships between neural regions sensitive to counterfactual comparisons, reward and punishment regions, and emotional experiences reported by participants than do disappointing wins. Further research is needed to support these hypotheses and untangle the functional and neural differences between disappointing wins and relieving losses.

Limitations and caveats

Though we had predicted that the effects of monetary comparisons would be much stronger for this study than monetary outcomes, we were surprised by the lack of a significant main effect of outcome in neural activity during the experienced outcome phase for this study. Upon further investigation, using a less stringent cluster threshold (300 voxels) resulted in the expected activity in the medial OFC for win outcomes and the left anterior insula for loss outcomes, consistent with existing research on neural responses to monetary wins and losses. We suspect that the weak signal in these regions is primarily due to the fact that comparisons were very salient, but it may also be due to the fact that our gambling paradigm differs from others used in the literature. In our paradigm, participants were informed of the ultimate valence of the outcome (win, loss) during the anticipation phase, only had two possible outcomes, and none of the outcomes were $0. The paradigm was purposefully designed to maximize the salience of obtained and unobtained outcomes by providing an obvious better outcome for each trial that participants would anticipate receiving. However, since participants learned they would win or lose money during the anticipation phase, this may have dampened activity in neural regions more typically found to respond to wins and losses. Given that anticipation and experience in the real world are often accompanied by the benefit of experience, prior knowledge, and environmental cues, we suggest that our paradigm accurately mimics a realistic psychological situation. Returning to a prior example, the anticipation of receiving a raise usually involves some knowledge about what is likely to occur, and we tend to experience the positive affect associated with the event before it comes to fruition (Loewenstein and Lerner, 2003). When we do finally get our raise, we tend to compare it to what we had expected to occur, and this comparison is usually accompanied by feelings of relief or disappointment. Future research, however, is needed to better understand neural differences in anticipation and experience in paradigms that maximize counterfactual thinking.

Conclusions

Overall, counterfactual comparisons are a prevalent and fundamental aspect of the human experience, and our study suggests that comparisons are what allow us to learn, emotionally recover, and make better decisions in our daily life. Activity in posterior parietal and dorsolateral prefrontal regions showed sensitivity to outcomes that produced mixed emotions due to counterfactual comparisons, and several regions suggested to process reward and punishment learning showed sensitivity to better and worse comparisons due to the existence of an unobtained outcome. Furthermore, activity in parietal and prefrontal regions sensitive to mixed outcomes correlated with activity in these reward regions, which further correlated with affect ratings of non-dominant emotions on mixed outcome trials. The emotionally conflicted feelings that occur with disappointment and relief, therefore, may result from comparisons carried out by the fronto-parietal network, which influences reward and punishment processing by shifting the focus to available comparisons to impact the ultimate emotional experience. In this way, our brain is able to capitalize on all available information and elicit accurate emotional responding to situations towards the goal of emotional recovery and driving better behavior in future endeavors.

Acknowledgments

The authors would like to thank John Cacioppo (University of Chicago) and Jeff Larsen (University of Tennessee) for helpful discussions regarding paradigm design. We also thank all of the members of the Norris Social Neuroscience Laboratory at Dartmouth College for help with data collection, especially Zachary Ingbretsen, Jad Bsat, and Stephanie Han.

References

Aggleton, J.P., 1992. The functional effects of amygdala lesions in humans: a comparison with findings from monkeys. In: Aggleton, J.P. (Ed.), The Amygdala: Neurobiological Aspects of Emotion, Memory and Mental Dysfunction. Wiley-Liss, New York, pp. 485–503.

Augustine, J.R., 1996. Circuitry and functional aspects of the insular lobe in primates including humans. Brain Res. Rev. 22, 229–244.

Behrmann, M., Geng, J.J., Shomstein, S., 2004. Parietal cortex and attention. Curr. Opin. Neurobiol. 14, 212–217.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B Methodol. 57, 289–300.

Breiter, H.C., Aharon, I., Kahneman, D., Dale, A., Shizgal, P., 2001. Functional imaging of neural responses to expectancy and experience of monetary gains and losses. Neuron 30, 619–639.

Bush, L.K., Hess, U., Wolfrod, G., 1993. Transformations for within-subjects designs: a Monte Carlo investigation. Psychol. Bull. 113 (3), 566–579.

Chua, H.F., Gonzalez, R., Taylor, S.F., Welsh, R.C., Liberzon, I., 2009. Decision-related loss: regret and disappointment. NeuroImage 47, 2031–2040.

Coricelli, G., Critchley, H.D., Joffily, M., O'Doherty, J.P., Sirigu, A., Dolan, R.J., 2005. Regret and its avoidance: a neuroimaging study of choice behavior. Nat. Neurosci. 8 (9), 1255–1262.

Cox, R.W., 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. 29, 162–173.

Creem-Regehr, S.H., 2009. Sensory-motor and cognitive functions of the human posterior parietal cortex involved in manual actions. Neurobiol. Learn. Mem. 91, 166–171.

Delgado, M.R., 2007. Reward-related responses in the human striatum. Ann. N. Y. Acad. Sci. 1104, 70–88.

Fink, G.R., Markowitsch, H.J., Reinkemeier, M., Bruckbauer, T., Kessler, J., Heiss, W.D., 1996. Cerebral representation of one's own past: neural networks involved in autobiographical memory. J. Neurosci. 16, 4275–4282.

Hirsch, J., Moreno, D.R., Kim, K.H., 2001. Interconnected large-scale systems for three fundamental cognitive tasks revealed by functional MRI. J. Cogn. Neurosci. 13 (3), 389–405.

Kahneman, D., Miller, D.T., 1986. Norm theory: comparing reality to its alternatives. Psychol. Rev. 93 (2), 136–152.

Kassam, K.S., Morewedge, C.K., Gilbert, D.T., Wilson, T.D., 2011. Winners love winning and losers love money. Psychol. Sci. 22, 602–606.

Larsen, J.T., McGraw, A.P., Cacioppo, J.T., 2001. Can people feel happy and sad at the same time? J. Pers. Soc. Psychol. 81 (4), 684–696.

Larsen, J.T., McGraw, A.P., Mellers, B.A., Cacioppo, J.T., 2004. The agony of victory and the thrill of defeat: mixed emotional reactions to disappointing wins and relieving losses. Psychol. Sci. 15 (5), 325–330.

Larsen, J.T., Norris, C.J., McGraw, A.P., Hawkley, L.C., Cacioppo, J.T., 2009. The evaluative space grid: a single-item measure of positivity and negativity. Cogn. Emotion 23 (3), 453–480.

Liu, X., Powell, D.K., Wang, H., Gold, B.T., Corbly, C.R., Joseph, J.E., 2007. Functional dissociation in frontal and striatal areas for processing of positive and negative reward information. J. Neurosci. 27 (17), 4587–4597.

Loewenstein, G., Lerner, J., 2003. The role of affect in decision making. In: Davidson, R.J., Goldsmith, H.H., Scherer, K.R. (Eds.), Handbook of Affective Science. Oxford University Press, Oxford, England, pp. 619–642.

Mars, R.B., Grol, M.J., 2007. Dorsolateral prefrontal cortex, working memory, and prospective coding for action. J. Neurosci. 27 (8), 1801–1802.

Mellers, B.A., McGraw, A.P., 2001. Anticipated emotions as guides to choice. Curr. Dir. Psychol. Sci. 10 (6), 210–214.

Mellers, B.A., Schwartz, A., Ho, K., Ritov, I., 1997. Decision affect theory: emotional reactions to the outcomes of risky options. Psychol. Sci. 8 (6), 423–429.

Mellers, B.A., Schwartz, A., Ritov, I., 1999. Emotion-based choice. J. Exp. Psychol. Gen. 128, 1–14.

Nicolle, A., Bach, D.R., Driver, J., Dolan, R.J., 2011. A role for the striatum in regret-related choice repetition. J. Cogn. Neurosci. 23 (4), 845–856.

Pinel, P., Dehaene, S., Riviere, D., LeBihan, D., 2001. Modulation of parietal activation by semantic distance in a number comparison task. NeuroImage 14 (5), 1013–1026.

Rozin, P., Royzman, E.B., 2001. Negativity bias, negativity dominance, and contagion. Pers. Soc. Psychol. Rev. 5 (4), 296–320.

Schimmack, U., 2001. Pleasure, displeasure, and mixed feelings: are semantic opposites mutually exclusive? Cogn. Emotion 15, 81–97.

Uddin, L.Q., Supekar, K., Amin, H., Rykhlevskaia, K., Nguyen, D., Greicius, M.D., Menon, V., 2010. Dissociable connectivity with the human angular gyrus and intraparietal sulcus: evidence from functional and structural connectivity. Cereb. Cortex 20 (11), 2636–2646.

Ursu, S., Carter, C.S., 2005. Outcome representations, counterfactual comparisons and the human orbitofrontal cortex: implications for neuroimaging studies of decision-making. Cogn. Brain Res. 23, 51–60.