2012-12-01

Prospect theory does not describe the feedback-related negativity value function.

Sambrook, TD

10.1111/j.1469-8986.2012.01482.x
Psychophysiology
Wiley

Psychophysiology, 49 (2012), 1533–1544. Wiley Periodicals, Inc. Printed in the USA.

DOI: 10.1111/j.1469-8986.2012.01482.x

Prospect theory does not describe the feedback-related negativity value function

THOMAS D. SAMBROOK, MATTHEW ROSER, AND JEREMY GOSLIN
Cognition Institute, School of Psychology, Plymouth University, Plymouth, UK

Abstract

Humans handle uncertainty poorly. Prospect theory accounts for this with a value function in which possible losses are overweighted compared to possible gains, and the marginal utility of rewards decreases with size. fMRI studies have explored the neural basis of this value function. A separate body of research claims that prediction errors are calculated by midbrain dopamine neurons. We investigated whether the prospect theoretic effects shown in behavioral and fMRI studies were present in midbrain prediction error coding by using the feedback-related negativity, an ERP component believed to reflect midbrain prediction errors. Participants' stated satisfaction with outcomes followed prospect theory but their feedback-related negativity did not, instead showing no effect of marginal utility and greater sensitivity to potential gains than losses.

Descriptors: Feedback negativity, Prospect theory, ERP, FRN, P300, Dopamine

Making effective decisions under choice requires people to construct mental representations of value. When goods on offer will be received with a certain probability only, a mathematically correct approach is to base decisions on expected value (EV), that is, the product of a good's value and the probability of obtaining it. Humans typically depart from this solution, and the science of behavioral economics attempts to account for this. The most successful account to date is prospect theory (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992). Prospect theory identifies three ways in which observed choices typically deviate from those dictated by expected value. First, following von Neumann and Morgenstern (1944), choices are based on utility, or the subjective value of the good. Subjective value shows diminishing growth with respect to objective value, producing the phenomenon of diminishing marginal utility. Second, this utility function is asymmetrically generated about a mathematically arbitrary but psychologically salient anchor point. The anchor point is typically zero, the utility function is typically steeper for outcomes less than zero, and this gives rise to the phenomenon of loss aversion; namely, that losses hurt more than equivalent gains please. Third, probabilities are distorted at the decision stage, such that small probabilities are inflated and large ones are diminished. It is the first two of the distortions described above, those that reify the zero anchor point, and are collectively known as the "value function," that are the subject of the current paper.

fMRI studies have identified a wide network of areas that appear responsive to value, and there is growing evidence that some of these may show a value function that is anchor point dependent. Pine et al. (2009) found that blood-oxygen-level-dependent (BOLD) activation in the striatum, anterior cingulate cortex (ACC), and ventral tegmental area showed effects consistent with diminishing marginal utility. Loss aversion has been studied by Tom, Fox, Trepel, and Poldrack (2007) who found that the striatum and ventromedial prefrontal cortex showed greater decreases in activity following losses than increases following gains, and that this effect correlated with individual behavioral differences in loss aversion. Loss-aversion brain behavior correlates have also been shown by de Martino, Kumaran, Holt, and Dolan (2009) and Seymour, Daw, Dayan, Singer, and Dolan (2007), specifically in the ventral striatum. In contrast, other studies have identified the amygdala and insula as sources of loss aversion (e.g., de Martino, Kumaran, Seymour, & Dolan, 2006; Knutson et al., 2008; Weber et al., 2007). The wider evidence for prospect theory effects in the brain is reviewed in Trepel, Fox, and Poldrack (2005).

A quite different body of work on neural representation of value concerns single cell recordings of midbrain dopamine neurons. These show phasic dopamine increases with unexpected reward and decreases with unexpected omission of reward (Bayer & Glimcher, 2005; Bayer, Lau, & Glimcher, 2007; Hollerman & Schultz, 1998; Schultz, Dayan, & Montague, 1997). They thus appear to code for the difference between expected and obtained value. Such prediction errors, or temporal difference errors, are known to be powerful tools in reinforcement learning (Sutton & Barto, 1998), and so there is reason to predict their representation in the brain.

It has been claimed by Holroyd and Coles (2002) that it is the phasic activity of midbrain dopaminergic neurons that gives rise to the electrophysiological (EEG) component known as feedback negativity, or feedback-related negativity (FRN). This component is characterized by a negative-going waveform at medial frontal sites in the interval 200–300 ms when unfavorable feedback is received. Holroyd and Coles argue that the FRN reflects the

Address correspondence to: Dr. Tom Sambrook, Cognition Institute, School of Psychology, Plymouth University, Plymouth, UK PL4 8AA. E-mail: tom.sambrook@plymouth.ac.uk

PSYCHOPHYSIOLOGY

transmission of a prediction error from the midbrain dopamine system to the ACC. As such, the amplitude of the FRN should be determined by the size of the prediction error. In keeping with the theory, subsequent research has shown that unexpected events produce greater FRNs (Bellebaum & Daum, 2008; Hajcak, Moser, Holroyd, & Simons, 2007; Holroyd & Coles, 2002; Potts, Martin, Burton, & Montague, 2006; San Martin, Manes, Hurtado, Isla, & Ibanez, 2010), as do outcomes of greater than expected magnitude (Bellebaum, Polezzi, & Daum, 2010; Kreussel et al., 2012; Marco-Pallares, Kramer, Strehl, Schroder, & Munte, 2010; Wu & Zhou, 2009, although see Hajcak, Moser, Holroyd, & Simons, 2006; Yeung & Sanfey, 2004).

Both the reinforcement learning theory of the FRN and prospect theory describe valuation. An important difference is that prospect theory describes valuation of outcomes *prior* to their resolution while the FRN purportedly represents valuation *after.* However, insofar as the FRN is believed to represent a prediction error, it must formally incorporate (and therefore be sensitive to) the prior valuation term. This is because prediction errors consist in the difference between obtained value and EV, and EV is given by the prior valuation of a prospective outcome (in conjunction with its probability). This means the FRN amplitude can be used to retrospectively infer the prior valuation placed on prospects. For example, if you flip a coin to either lose a dollar or lose nothing and you wind up losing the dollar, this outcome constitutes a prediction error of -$0.50. However, if your prior valuation of the loss of a dollar was in fact -$2 (as claimed by prospect theory), then the outcome constitutes a prediction error of -$1. Consequently, where FRNs differ despite no difference in the objective value of prediction errors, we can infer distorted prior valuation.

While the existence in the brain of a biased estimate of EV is implied by the ubiquity of anchor point effects on behavior, it does not follow that this is used in the generation of prediction errors. The discussion of fMRI findings above suggests the possibility of multiple representations of EV, and while anchor point effects are certainly present at the generation of behavior, it is unclear at what point they are introduced neurally. It might be relatively late, at the point of deciding between alternative actions. Alternatively, anchor point biases might be applied at the very outset of EV computation, prior to any involvement of systems concerned with behavioral output. Since, as noted above, the FRN should be sensitive to any anchor point effects that are present in the EV term it uses to construct prediction errors, it can be used to investigate when, in the route from valuation to behavior, these anchor point effects are applied. Because the FRN concerns evaluation of feedback, rather than choice between actions, if it reveals anchor point effects, then this suggests that these are widespread in the brain's representation of EV. If they are absent, it might suggest that they are introduced later and closer to the point of action selection.

Behaviorally, anchor point effects have been demonstrated in nonhuman primates (Chen, Lakshminarayanan, & Santos, 2006; Lakshminarayanan, Chen, & Santos, 2011) and birds (Marsh & Kacelnik, 2002), suggesting they may have an extracortical source. However, to date, single cell studies have not been revealing. Regarding diminishing marginal utility, Bayer and Glimcher (2005) found a simple linear relationship between positive prediction error and dopamine release, while others (e.g., Morris, Arkadir, Nevet, Vaadia, & Bergman, 2004; Satoh, Nakai, Sato, & Kimura, 2003) conclude that the relationship was monotonic. There is no single cell work on loss aversion that we are aware of.

In the following experiments, participants undertook a number of gambles that could lead to a good or a bad outcome with respect to the gamble's EV. The monetary difference between the good and the bad outcome was held constant in each experiment. This was important because it led to the null hypothesis that differences in the feedback event-related potential (ERP) following good and bad outcomes would remain constant regardless of other experimental factors. This null hypothesis is based on the assumption that the FRN uses an EV term that is unaffected by the zero anchor point and thus codes obtained value and expected value objectively.

In Experiment 1, we manipulated the factor of domain, that is, whether the sum gambled for was gained or lost from participants' bankrolls. Gambles either occurred in the gain domain so that participants won 10 units or won nothing, or they occurred in the loss domain so that participants lost 10 units or lost nothing. The phenomenon of loss aversion is thus a domain effect. While gambles in the gain domain are clearly preferable to those in the loss domain, the difference between the good and bad outcomes is the same in both gambles, objectively speaking. Loss aversion, in contrast, predicts that the perceived difference will be greater in the loss domain. If this applies to the FRN, then a greater difference should be seen between the waveforms for good and bad outcomes in the loss (−10/0) domain gamble than the win (0/10) domain gamble.

In Experiment 2, we manipulated the factor of zero proximity, that is, whether the sum gambled for lay close to or far from zero. Gambles were either zero proximal such that participants won 10 units or nothing or they were zero distal such that participants won 100 units or 90. The phenomenon of diminishing marginal utility is thus a zero-proximity effect. Gambles that are zero distal are clearly preferable to those that are zero proximal; however, once again, the difference between the good and bad outcomes is the same in both gambles. Diminishing marginal utility predicts, however, that the perceived difference will be greater for zero-proximal gambles. If this applies to the FRN, then a greater difference should be seen between the waveforms for good and bad outcomes in the zero-proximal (0/10) gamble than the zero-distal (90/100) gamble. The generation of prospect theory effects from the underlying value function is shown in Figure 1.

Regarding the zero-proximity manipulation, we know of no FRN studies that have studied this. Regarding domain, Holroyd, Larsen, and Cohen (2004) found no evidence of this effect when gambles of equal stakes were offered either side of zero. However, looking for loss aversion was not the object of their study, and these authors blocked the gambles in the gain and loss domains separately. Given the importance of framing effects in prospect theory (Tversky & Kahneman, 1981), we would expect this to strongly reduce or eliminate the salience of the zero anchor point and attenuate loss aversion. Kreussel et al. (2012) recently made a comparison of gambles in the loss and gain domains as part of a wider study of the determinants of the FRN. These authors found no effect of domain unless time (first vs. second half of the experiment) was included as a factor, in which case FRN showed increasingly greater sensitivity to gambles in the gain domain. This is an intriguing result, contrary to prospect theory. Nevertheless, the requirement of time as a mediating variable, and the fact that a further two other variables were manipulated in Kreussel et al.'s experiment, increasing the danger of component overlap and also perhaps diminishing the salience of the zero point, suggests further exploration is appropriate. In addition, we also report P300 amplitudes since this component has been shown to be sensitive to the determinants of EV (e.g., Wu & Zhou, 2009) and the stimulus-preceding negativity (SPN), a slow negative wave that precedes stimuli containing information and which is believed to be an index of attention (Brunia & van Boxtel, 2012).

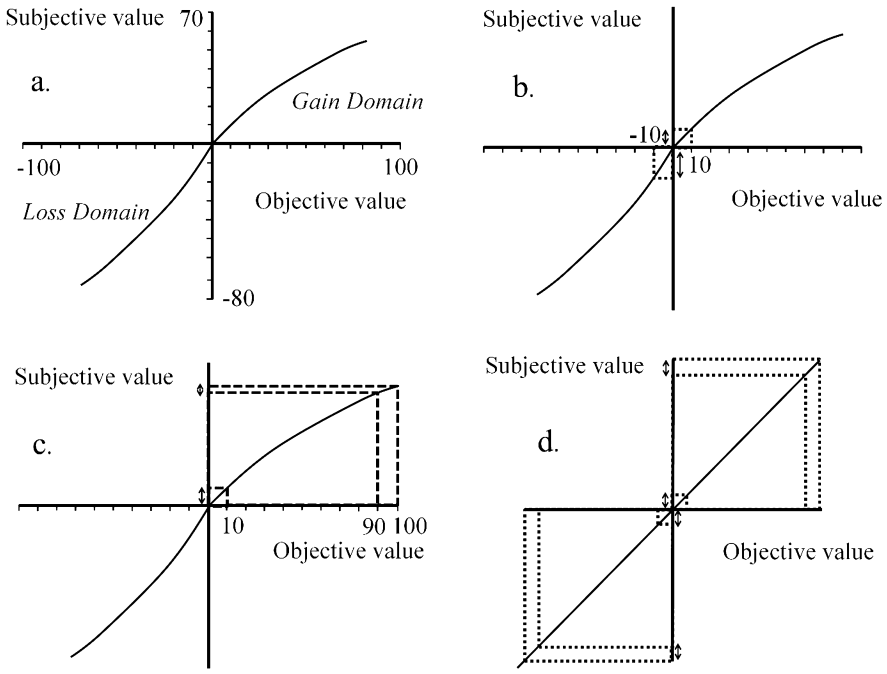

Figure 1. a: The prospect theory value function is curved and is steeper for losses. b: A loss of 10 units has more subjective impact than a gain of 10 units as shown by the double-headed arrows. c: A gain of 10 units over an initial nothing has greater subjective impact than a gain of 10 units over an initial 90. d: A value function where there is no zero anchor point: direction and distance from zero have no effect on subjective value.

Experiment 1

Method

Participants. Seventy-two undergraduates (3 left-handed, 10 male) participated for course credit and the opportunity to win a small sum of money. Data from six participants were rejected (three for equipment failure, two for artifacts on over 75% of trials, one for blinking selectively in response to bad feedback).

Task. Since money gained and lost on the task was predetermined to be zero, it was necessary to provide participants with a starting bankroll to avoid the possibility of their having negative winnings during the game. To encourage a sense of ownership of this money, participants "won" it in a reaction time task unrelated to the main experiment. While winnings were ostensibly related to reaction times, all participants in fact won £2.90 (approximately $4.50).

The primary task was presented on a computer using E-Prime software and is depicted in Figure 2. Participants were shown a graphic depicting four squares, each containing a symbol. Two of the squares were highlighted with a thick border. Participants were told that on each occasion one of the highlighted symbols would be good and the other bad. They selected one of the two available symbols using keys corresponding to the icon's position on the screen.

Prior to learning whether their choice was correct, participants were instructed to choose which domain to play in, that is, "for a win" or "to avoid loss." This was accomplished with a graphic showing these phrases either side of the screen. On the first of any set of four trials, these phrases appeared in duplicate. Selecting an option resulted in one fewer of the corresponding phrase appearing on the next trial. When both copies of a phrase had disappeared, choosing those stakes was not permissible. This ensured that in a block of four trials participants played twice for a win and twice to avoid loss in an order of their choosing. Participants were made to actively choose the domain to increase the salience of this key manipulation and to encourage its retention up to the point of feedback. For the same reason, participants chose it after they selected the symbol that ostensibly determined whether the outcome would be good or bad.

An auditory confirmation of participants' choice of domain lasted 1,250 ms, immediately followed by a fixation cross presented for 500–600 ms. Feedback was then presented for 5 s in the form of a pink or yellow circle, denoting whether the good or bad symbol had been chosen (the color representing the good symbol was counterbalanced across participants). The circle appeared in the center of the screen against a white background and occupied approximately 3° of visual angle horizontally and vertically. Finally a summary window appeared for 1 s detailing that trial's winnings and money earned so far. "To win trials" resulted in a win of £0.10 or nothing; "to avoid loss" trials resulted in a loss of £0.10 or nothing. Outcomes were pseudorandomly determined to be good or bad half the time for each domain. The position of the four symbols, which were highlighted, was randomly determined on each trial. While only two symbols were necessary to implement the experiment, extra symbols were used to obscure the random nature of the task and to encourage the formation of strategies that would lead to greater interest in feedback (Yeung, Holroyd, & Cohen, 2005).

To check that participants were paying attention both to the domain played in and the valence of the feedback, after approximately 10% of trials (randomly determined) they were prompted to press a key to indicate what domain the preceding trial had been played in and what its outcome had been. A 2-s time limit was placed on each response. Participants' bankrolls were increased or decreased £0.10 for right or wrong answers. Total earnings for the experiment averaged £4.62 per person.

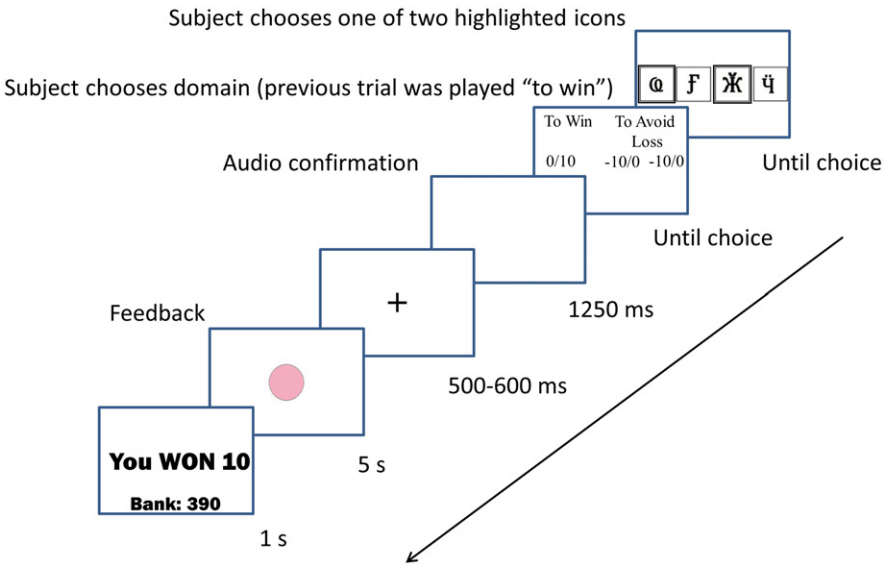

Figure 2. Sequence of stimuli in Experiment 1. Participants chose a highlighted icon, selected a domain to play in, received audio confirmation of their choice, fixated, received color-coded feedback on whether a good or bad outcome had been achieved, and were shown a summary of winnings so far.

Participants played through 192 trials in four blocks separated by 30-s breaks. At the experiment's conclusion, the satisfaction of the participant with the various outcomes was measured. Participants were shown a graphic depicting a 25-cm line and told one end represented "very unhappy" and the other "very happy." They were asked to position markers on the line to indicate how, on average, the good and bad outcomes for each of the two domains had made them feel. These were converted to numerical scores. The order in which the markers were placed was counterbalanced across participants.

EEG recording and analysis. EEG data were collected from 11 actively amplified Ag/AgCl electrodes (actiCAP, Brain Products, Gilching, Germany) mounted on an elastic cap. The electrodes were Fz, FCz, Cz, CPz, Pz, F3, F4, FC3, FC4, FP1, FP2. Since the primary focus of this study was the FRN, which is reliably maximal at medial sites, the lateral sites were included merely to allow a preliminary qualitative check that the component identified as the FRN was not greater laterally than medially. Electrodes were referenced to the left mastoid and rereferenced offline to the average of left and right mastoid activity. Vertical eye movement was monitored by electrodes FP1 and FP2 and a right suborbital electrode, and horizontal eye movement was monitored using an electrode on the right external canthus. Electrode impedances were kept below 20 kΩ. EEGs were amplified using a BrainAmp amplifier (Brain Products), continuously sampled at 500 Hz, and filtered offline with a band-pass filter from 0.1 to 30 Hz. ERPs were computed by averaging artifact-free EEGs (> 90% trials). EEGs were rejected if FP1, FP2, or ocular electrodes showed a voltage change exceeding 100 μv/100 ms or if any midline site showed a voltage change exceeding 20 μv/ms.

For the FRN and P300, EEGs were time-locked to 200 ms before the onset of the feedback to 700 ms afterward, and then were baseline-corrected using the period −100 to 0 ms. A problem that has hampered FRN research is lack of consistency in how the component is measured, which can result in the same waveform producing markedly different measures of FRN amplitude depending on the measure. Each method has its relative merits, as discussed by Wu and Zhou (2009). To increase the robustness of our conclusions and to facilitate comparison with the existing and future literature we quantified the FRN using three representative methods: (1) A peak-to-average-peak measure was found by locating three peaks of alternating polarity: a positive peak from 150 to 280 ms, a following negativity from 200 to 350 ms, and a following positivity from 280 to 500 ms. FRN was calculated as the difference of the negative peak and the average of the adjacent positive peaks. A peak-to-average-peak measure was preferred to a peak-to-peak measure that used only the preceding positivity because grand average waveforms revealed that bad outcomes were associated with a late positivity at ~425 ms that was not attributable to the P300, which showed positivity for good outcomes. The same peak-to-average-peak measure of the FRN was used by Gu et al. (2011), Oliveira, McDonald and Goodman (2007) and Yeung and Sanfey (2004), all of whom present waveforms in which a late positive deflection is similarly associated with bad outcomes. (2) A mean voltage measure was taken in the interval 200–300 ms. For this measure, the data were high-pass filtered at 2 Hz in order to remove slow wave activity associated with the P300 (Wu & Zhou, 2009). (3) A difference wave measure was taken by finding the peak of the difference wave of good and bad outcomes in the interval 200–300 ms.

The P300 was measured as the peak positivity in the interval 275–500 ms. For the SPN, EEGs were time-locked to 1,250 ms prior to the onset of the feedback (500–600 ms following selection of domain) and baseline-corrected using the period −1,250 to −750 ms. SPN was measured as the mean voltage in the interval −200 to 0 ms preceding feedback onset (Kotani et al., 2003). When analyzing the data, Greenhouse-Geisser corrections were applied where appropriate.

Results

Behavioral results. The fact that feedback was pseudorandomly assigned precludes meaningful behavioral analysis of the primary

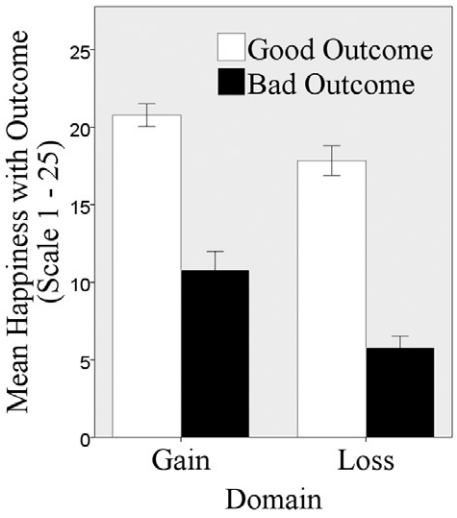

Figure 3. Participants' happiness with each of the four possible Domain × Valence outcomes as measured by self-report at the end of the experiment.

task. Performance on the randomly delivered probe questions showed that participants' attention was high (errors recalling domain < 5%, outcome < 4%).

The affective ratings that participants gave at the end of the experiment for the four outcomes are shown in Figure 3. A 2 × 2 analysis of variance (ANOVA) using the factors of valence and domain revealed greater happiness for good outcomes, $F(1,65) = 441.67$, $p < .001$, $\eta_p^2 = .87$, for outcomes in the gain domain, $F(1,65) = 89.78$, $p < .001$, $\eta_p^2 = .58$, with an interaction between these factors, $F(1,65) = 9.48$, $p \leq .01$, $\eta_p^2 = .13$, that revealed the valence effect to be stronger in the loss domain, as predicted by prospect theory. Further pairwise t tests revealed that there were significant differences ($p < .05$) between each of the individual components of this interaction.

Electrophysiological results. Feedback-related ERP activity typically consists in a frontocentral FRN between 200 and ~300 ms that is more negative for bad outcomes, and a parietal P300 from ~270 ms onwards that is more positive for good outcomes. The bad–good outcome difference term should thus be negative for both components, and this can be seen in both the ERP waveform and topographical maps in Figure 4.

A one-way ANOVA testing the effect of midline electrode site (Fz to Pz) on difference wave peak negativity in the interval 200–300 ms revealed a significant effect of site, $F(4,260) = 8.67$, $p < .001$, $\eta_p^2 = .12$, with the difference wave peak greatest at FCz (4.16 μv) and smallest at Pz (3.97 μv). Consequently, FCz was chosen as the site at which the FRN was measured.

Figure 5 shows the waveforms elicited by good and bad outcomes in the two domains. As described earlier, three FRN measures were employed. A peak-to-average-peak analysis revealed a significant effect of outcome valence, with FRN greater for bad outcomes, $F(1,65) = 4.17$, $p < .05$, $\eta_p^2 = .06$, and an interaction between valence and domain, $F(1,65) = 8.19$, $p < .01$, $\eta_p^2 = .11$. Simple effects analysis showed that the valence effect was present in the gain domain, $t(65) = 3.14$, $p < .01$, but not the loss domain, $t(65) = .24$, $p > .05$. Further pairwise t tests of this interaction revealed that the only significant differences ($p < .05$) were as a result of the FRN being significantly higher in the gain/bad condition than all other conditions. There was no main effect of domain, $F(1,65) = 2.05$, $p > .05$.

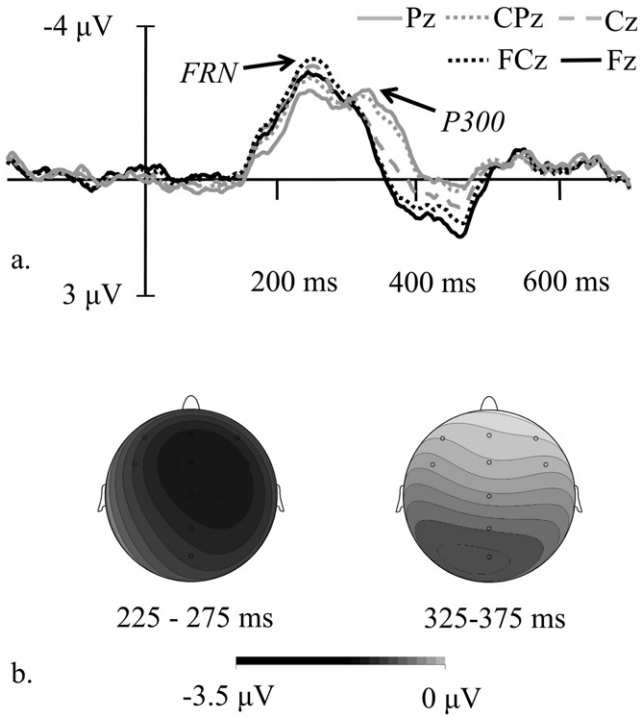

Figure 4. The effect of feedback valence (bad–good) in Experiment 1 shown as an ERP waveform (a) and a scalp topography (b).

A mean voltage measure of FRN revealed a similar pattern of results: significantly greater FRN for bad outcomes, $F(1,65) = 43.45$, $p < .001$, $\eta_p^2 = .40$, and a significant interaction between valence and domain, $F(1,65) = 17.11$, $p < .001$, $\eta_p^2 = .18$. However, simple effects analysis showed that the valence effect was present for both gain, $t(65) = 7.43$, $p < .001$, and loss, $t(65) = 4.31$, $p < .001$, domains. There was no main effect of

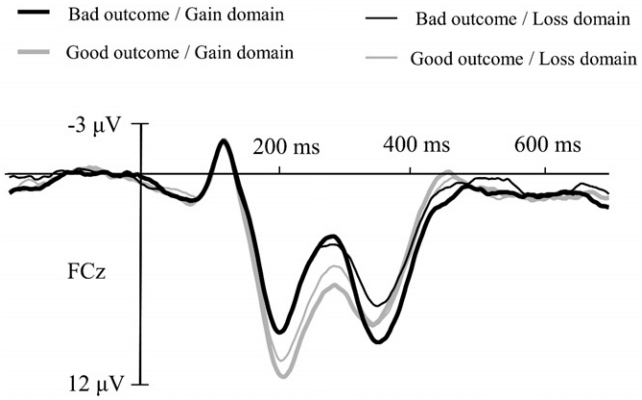

Figure 5. FRN in Experiment 1. The figure shows grand average ERPs following feedback on whether a good or bad outcome had been achieved from a monetary gamble. Good outcomes are shown by light lines, bad outcomes by dark lines. Gambles in the gain domain are shown by thick lines, gambles in the loss domain by thin lines. Feedback is at 0 ms. Electrode site is FCz.

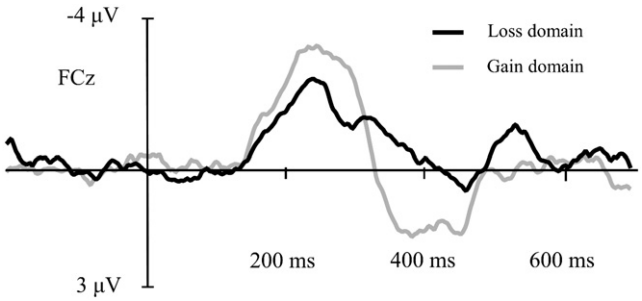

Figure 6. Difference waves created from waveforms shown in Figure 5. The good outcome waveform is subtracted from the bad outcome waveform for each domain separately.

domain, $F(1,65) = 3.41$, $p > .05$. Further pairwise t tests revealed that there were significant differences ($p < .05$) between each of the individual components of this interaction with the exception of the gain/bad and loss/bad comparison.

Figure 6 shows bad–good outcome difference waves for the two domains. The difference wave peak in the gain domain was significantly greater, $t(65) = 3.12$, $p < .01$, thus supporting the interaction between domain and valence shown in the previous measures.

Following Kreussel et al. (2012), we examined the effect of time by splitting the data into two halves. Three participants were removed because trials fell below the recommended minimum of 20 (Marco-Pallares, Cucurell, Munte, Strien, & Rodriguez-Fornells, 2011). Using a peak-to-average-peak analysis, the three-way Domain × Valence × Time interaction fell just short of significance, $F(1,62) = 3.36$, $p < .10$. We explored this further by checking whether the Valence × Domain interaction shown previously was present in both halves of the experiment. There was no interaction in the first half, $F(1,62) = 1.26$, $p > .05$, but a strong interaction in the second, $F(1,62) = 12.53$, $p < .001$, $\eta_p^2 = .17$. Using a mean voltage measure, no significant Time × Valence × Domain interaction was found ($F < 1$).

A one-way ANOVA testing the effect of midline electrode site (Fz to Pz) on the positivity of the waveform (all conditions averaged) in the interval 275–500 ms revealed a significant effect of site, $F(4,260) = 8.67$ $p < .001$, $\eta_p^2 = .12$, with the peak positivity greatest at Pz (10.52 μv) and smallest at Fz (7.40 μv). Consequently, Pz was chosen as the site at which the P300 was measured. The P300 is shown in Figure 7. An analysis of the P300 revealed that its amplitude was significantly greater for good outcomes, $F(1,65) = 25.94$, $p < .001$, $\eta_p^2 = .21$, and for outcomes in the gain domain, $F(1,65) = 8.93$, $p < .01$, $\eta_p^2 = .12$, but there was no interaction between valence and domain, $F < .01$.

A one-way ANOVA testing the effect of midline electrode site (Fz to Pz) on the mean negativity of the waveform (all conditions averaged) in the interval −200 to 0 ms prefeedback revealed a significant effect of site, $F(4,260) = 16.34$, $p < .001$, $\eta_p^2 = .20$, with greatest negativity at Cz (−7.48 μv) and least at Fz (−5.28 μv). Consequently, Cz was chosen as the site at which the SPN was measured. Results revealed that the SPN was significantly greater preceding loss domain feedback, $t(65) = 2.46$, $p < .05$.

Discussion

The stated satisfaction of participants with outcomes showed strong loss aversion, but there was no evidence of this in the FRN. Instead, the reverse effect was found: increased sensitivity for gains. The P300 was elevated for outcomes that were good and for those in the gain domain. The SPN, an index of anticipatory attention, was greater for loss domain gambles, in keeping with the sensitivity shown to the loss domain in participants' stated satisfaction.

Experiment 2

Method

Participants. Fifty-nine undergraduates (5 left-handed, 24 males) participated for course credit and the opportunity to win a small sum of money. No participants were subsequently rejected.

Task. The task was essentially similar in structure to Experiment 1 with the following differences. As domain and marginal utility terms are orthogonal in Kahneman and Tversky's (1979) mathematical realization of prospect theory, the gambles in this experiment were presented in a single domain. The gain domain was chosen as the effect of valence was found to be stronger in the gain than the loss domain in Experiment 1, while also being the most commonly studied domain within the literature. Since participants could not incur negative winnings, the reaction time task from Experiment 1 was omitted. In the main experimental task, in order to provide winnings that were either close or far from the zero anchor point, and at the same time keep participant payments manageable, the game was played for points. On a given trial, participants could earn 0, 10, 90, or 100 points. These were later converted to cash at the rate of £0.01 per 25 points. All other events on a trial resembled Experiment 1 except that instead of choosing to play "to win" or "to avoid loss," participants chose gambles that were either close to or far from zero, namely, "nothing or ten" or "ninety or a hundred." These choices were displayed as "0/10" and "90/100." Total earnings for the experiment averaged £4.82 per person.

Results

Behavioral results. Performance on the randomly delivered probe questions showed that participants' attention was high (errors

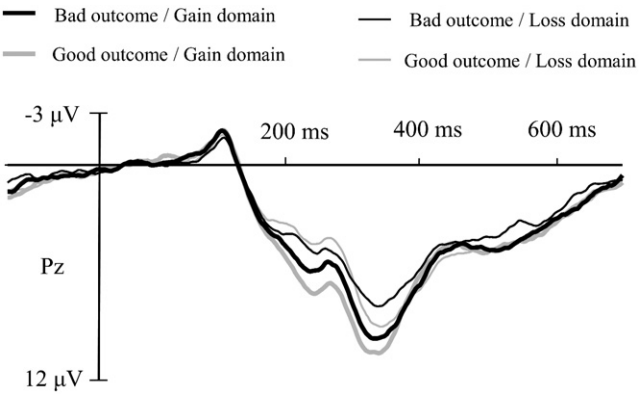

Figure 7. P300 in Experiment 1. The figure shows grand average ERPs following feedback on whether a good or bad outcome had been achieved from a monetary gamble. Good outcomes are shown by light lines, bad outcomes by dark lines. Gambles in the gain domain are shown by thick lines, gambles in the loss domain by thin lines. Feedback is at 0 ms. Electrode site is Pz.

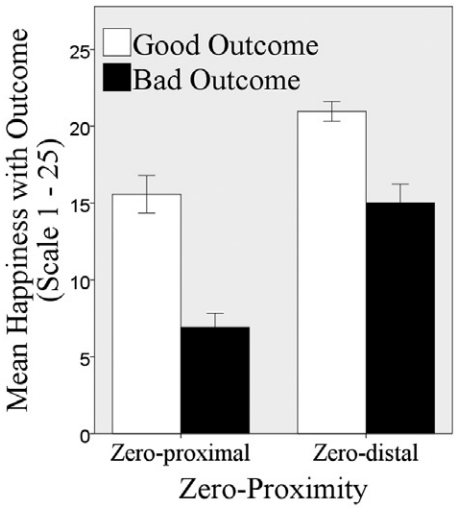

Figure 8. Participants' happiness with each of the four possible Zero-Proximity × Valence outcomes as measured by self-report at the end of the experiment.

recalling zero proximity < 5%, outcome < 3%). The affective ratings that participants gave at the end of the experiment for the four outcomes are shown in Figure 8. To test for an effect of diminishing marginal utility, a 2 × 2 ANOVA of zero proximity and outcome valence was performed. This revealed greater happiness for good outcomes, $F(1,58) = 341.52$, $p < .001$, $\eta_p^2 = .86$, for zero-distal outcomes, $F(1,58) = 131.85$, $p < .001$, $\eta_p^2 = .69$, and an interaction, $F(1,58) = 13.21$, $p < .001$, $\eta_p^2 = .19$, such that the valence effect was stronger for the zero-proximal outcomes than the zero-distal ones, as predicted by prospect theory. Further examination using pairwise *t* tests revealed that there were significant differences ($p < .05$) between each of the individual components of this interaction with the exception of the zero-proximal/good versus the zero-distal/bad outcome.

Electrophysiological results. Figure 9 shows bad–good outcome voltage differences. It shows the same progression of FRN to P300 seen in Experiment 1. The site at which FRN was maximal (using the negative peak of the difference wave at 200–300 ms) was established using a one-way ANOVA using the five midline electrodes. While this did not produce a significant effect of site, ($F < 1$), the difference wave peak was greatest at FCz (4.14 μv) and smallest at Pz (3.96), and thus FCz was chosen as the site at which the FRN was measured.

Figure 10 shows the ERPs elicited by good and bad outcomes for zero-proximal (0/10) and zero-distal (90/100) gambles. A peak-to-average-peak analysis revealed a significant effect of valence, with FRN greater for bad outcomes, $F(1,58) = 9.21$, $p < .01$, $\eta_p^2 = .14$. There was no effect of zero proximity, $F < .1$, nor was there a significant interaction between zero proximity and valence, $F < .1$ A mean voltage measure of FRN revealed the same pattern of results: a significantly greater FRN for bad than good outcomes, $F(1,58) = 42.51$, $p < .001$, $\eta_p^2 = .42$, but no main effect of zero proximity, $F(1,58) = 2.54$, $p > .05$, or interaction with valence, $F < .1$. Figure 11 shows bad–good outcome difference waves, with the peak of difference wave measure of FRN in agreement with the previous analyses in showing no effect of zero proximity, $t < 1$.

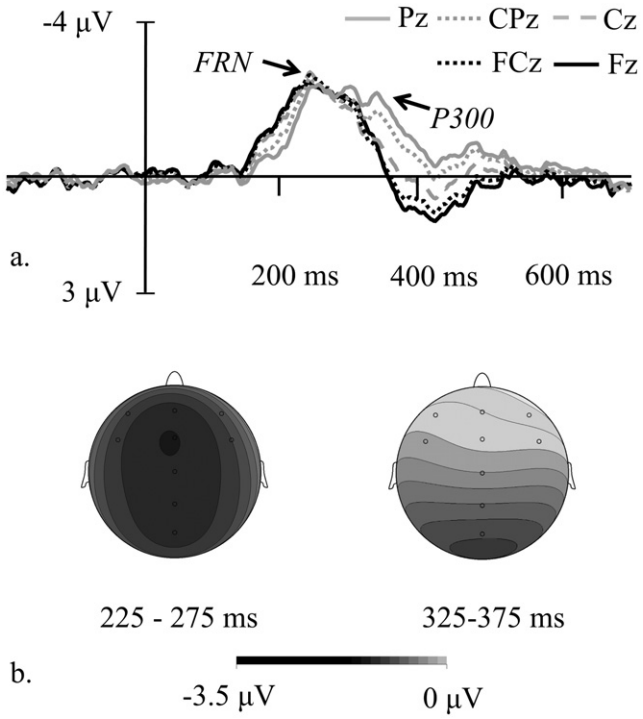

Figure 9. The effect of feedback valence (bad–good) in Experiment 2 shown as an ERP waveform (a) and a scalp topography (b).

A one-way ANOVA testing the effect of midline electrode site (Fz to Pz) on the positivity of the waveform (all conditions averaged) in the interval 275–500 ms revealed a significant effect of site, $F(4,232) = 34.49$, $p < .001$, $\eta_p^2 = .37$, with the peak positivity greatest at Pz (8.72 μv) and smallest at Fz (5.32 μv). Consequently, Pz was chosen as the site at which the P300 was measured. An analysis of the P300, shown in Figure 12, revealed that the amplitude of this component was significantly greater for good outcomes, $F(1,58) = 18.12$, $p < .001$, $\eta_p^2 = .24$, but that there was no effect of zero proximity, $F < 1$, or interaction with valence, $F < 1$.

A one-way ANOVA testing the effect of midline electrode site (Fz to Pz) on the mean negativity of the waveform (all conditions

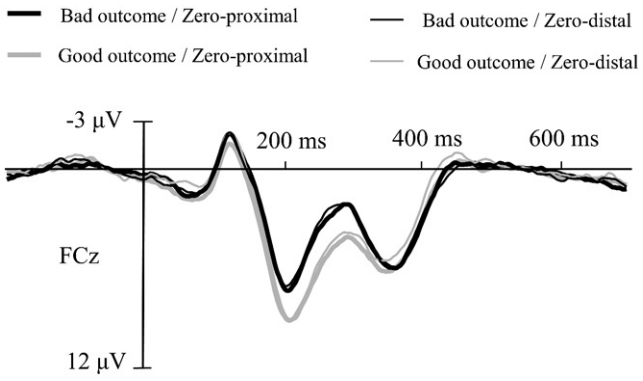

Figure 10. FRN in Experiment 2. The figure shows grand average ERPs following feedback on whether a good or bad outcome had been achieved from a monetary gamble. Good outcomes are shown by light lines, bad outcomes by dark lines. Zero-proximal gambles are shown by thick lines, zero-distal gambles by thin lines. Feedback is at 0 ms. Electrode site is FCz.

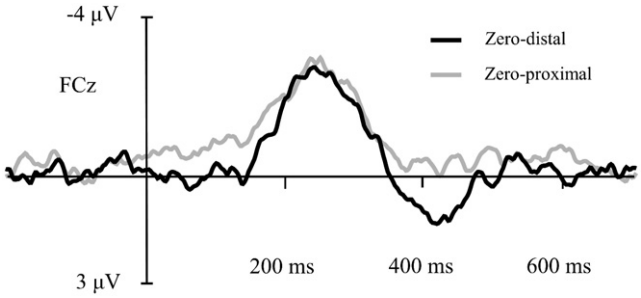

Figure 11. Difference waves created from waveforms shown in Figure 10. The good outcome waveform is subtracted from the bad outcome waveform for the zero-proximal and zero-distal gambles separately.

averaged) in the interval −200 to 0 ms prefeedback revealed a significant effect of site, $F(4,232) = 10.09$, $p < .001$, $\eta_p^2 = .15$, with greatest negativity at Cz (−6.03 μv) and least at Pz (−4.74 μv). Consequently, Cz was chosen as the site at which the SPN was measured. The SPN showed no significant difference between zero-proximal and zero-distal gambles, $t < 1$.

Discussion

As with Experiment 1, participants' stated satisfaction with outcomes followed prospect theory, showing greater sensitivity to zero-proximal outcomes. However, no such effect was observed in the FRN, which was merely responsive to outcome valence. The P300 was likewise sensitive only to valence. There was no effect of zero proximity on the SPN.

General Discussion

Encoding of value is believed to start at the midbrain. It finds its final expression in behavior and conscious appraisal. At some point in this process, people typically succumb to two anchor point–dependent biases in value coding, sensitivity to the domain (win or loss), and sensitivity to zero proximity (zero proximal or zero distal). When people make judgments of the value of prospects, that is, of their expected value, they appear to make them with respect to these biased values.

The rationale of the experiments presented here was to see if these biases of expected value were used in the generation of the FRN. The FRN is believed to represent a prediction error, that is, a comparison of obtained value with EV. If the EV term used in the generation of the FRN showed anchor point effects, this would suggest that these were an integral feature of valuation in the brain, not a bias introduced only at the point at which behavior was implemented.

The biases, evident in participants' affective ratings, were not present in the FRN, suggesting anchor point effects may not be integral to EV representation. Our data on diminishing marginal utility in Experiment 2 are consistent with Bayer and Glimcher's (2005) demonstration of a linear coding of prediction error magnitude in single dopamine cells (at least for positive errors). Our confidence in this null result is bolstered by the large sample size and the fact that essentially the same procedure was capable of eliciting an effect in Experiment 1. A post hoc power estimate can be usefully applied here. In Experiment 1, the observed power of the design in finding the significant interaction between domain and valence with an effect size of $\eta_p^2 = .11$ was .85. While Experiment 2 had very slightly fewer participants, the close design similarities suggest a similar power value, but in this case the interaction between valence and zero proximity produced a p value of .8, and so we are reasonably confident that these factors do not interact.

In Experiment 1, in contrast to loss aversion, participants' FRNs showed greater sensitivity to gains. There is, to our knowledge, no single cell study of domain effect to which we can compare this result. However, since the FRN is believed to arise when midbrain prediction errors reach either the cortex (Holroyd & Coles, 2002) or the basal ganglia (see below), it would be unwise to assume that this observed gain sensitivity is necessarily present in midbrain coding. Rather, it may arise from modulation of an initially unbiased midbrain prediction signal. In either event, however, this study suggests a value function for the FRN that is quite different from the one proposed by prospect theory as shown in Figure 1a. While a graded, parametric manipulation of the factors of domain and zero proximity is ultimately required to map the FRN value function, evidence from this study provisionally suggests a linear rather than curved function (Experiment 2), which is steeper for gambles in the gain domain than in the loss domain (Experiment 1).

An alternative explanation for the observed Domain × Valence interaction in Experiment 1 might arise from the fact that the gain/good outcome is dissimilar to the other three outcomes in that there is no aspect of that outcome which is negative. It might thus constitute a qualitatively different outcome that drives the interaction and supposed greater sensitivity to gains. However, t tests revealed that the gain/good outcome was not an outlier. Under a mean voltage measure of the FRN, both the gain/good and gain/bad outcomes showed significant differences from other conditions, while under a peak-to-average-peak measure, it was the gain/bad condition that was an outlier compared to the other conditions.

In order to integrate our findings with the existing fMRI literature on prospect theory effects, it is necessary to consider the neural generator of the FRN. The ACC has traditionally been believed to be the source of the FRN. However, recent evidence has suggested that at least some portion of the FRN may arise in the striatum. Using both fMRI and ERPs, Carlson, Foti, Mujica-Parodi, Harmon-Jones, and Hajcak (2011) found that FRN correlated with

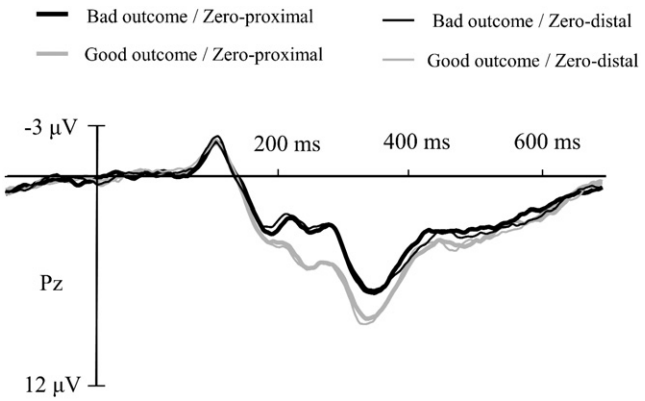

Figure 12. P300 in Experiment 2. The figure shows grand average ERPs following feedback on whether a good or bad outcome had been achieved from a monetary gamble. Good outcomes are shown by light lines, bad outcomes by dark. Zero-proximal gambles are shown by thick lines, zero-distal gambles by thin lines. Feedback is at 0 ms. Electrode site is Pz.

BOLD activation in the ventral striatum, dorsal striatum, and medial prefrontal cortex. However, principal components analysis made in that study and another conducted by Foti, Weinberg, Dien, and Hajcak (2011b) identified the dorsal striatum as the most probable generator of the FRN. This was in agreement with the conclusions of Martin, Potts, Burton, and Montague (2009). The issue remains unresolved (Cohen, Cavanagh, & Slagter, 2011; Foti, Weinberg, Dien, & Hajcak, 2011a); however, the frequency with which the striatum is found to be activated by both rewards and nonrewards (see Liu, Hairston, Schrier, & Fan, 2011, for a meta-analysis) justifies its consideration as a contributor to the FRN and raises some possibilities, which we consider below.

Our domain effect stands in direct contrast to the finding of Tom et al.'s (2007) fMRI study. This showed that the striatum responded to outcomes in both gain and loss domains but with increased sensitivity in the loss domain. These authors suggested that the striatum was a source of value coding with an inbuilt loss aversion parameter. Others have proposed a more componential basis for value coding. For example, both Yacubian et al. (2006) and Knutson, Adams, Fong, and Hommer (2001) found the ventral striatum to be responsive only to gains, while losses and loss aversion were associated with activity in the amygdala (de Martino et al., 2006; Weber et al., 2007; Yacubian et al., 2006) or insula (Knutson et al., 2008).

A resolution of these conflicting findings that would also be consistent with those of the present study concerns the differential roles that might be played by dopamine and serotonin. Opponency between dopamine and serotonin is well established (e.g., Daw, Kakade, & Dayan, 2002). In the striatum, there is evidence of a serotonin-dopamine gradient along a caudal-rostral axis (Brown & Molliver, 2000; Heidbreder, Hedou, & Feldon, 1999). Intriguingly, Seymour et al. (2007) found that, within the striatum, anterior regions showed relative selectivity for rewards, and posterior regions for losses. This accords with a number of recent studies, which have suggested that these two neuromodulators contribute differentially to coding for outcomes less than or greater than the zero anchor point. For example, Zhong et al. (2009) have shown that genetic variation in tonic dopamine and serotonin levels modifies risk seeking in gain and loss domains, respectively. Pessiglione, Seymour, Flandin, Dolan, and Frith (2006) showed that dopamine agonists affected choices in the gain domain (both neurally and behaviorally) but not the loss domain. Campbell-Meiklejohn et al. (2011) showed that serotonin but not dopamine promoted "chasing behavior," that is, gambles undertaken by participants who consider themselves to currently hold assets below an anchor point (typically what they started a session with).

Given that gambling is a recent, purely cultural phenomenon, it may initially seem implausible to suggest differential action of neuromodulators for gambles either side of the zero anchor point. It appears more credible, however, when one considers that the anchor point effectively frames all outcomes below it as a form of punishment and all outcomes above it as a form of reward. Thus, gambling may be rather easily mapped onto a much older functional division. Consequently, it is possible that gambles with an EV higher than zero might harness a dopaminergic system devoted to detecting rewarding/appetitive outcomes (and their omission), and gambles with negative EV might harness a serotoninergic system devoted to detecting punishment/aversive outcomes (and their avoidance). There is no evidence of serotoninergic neurons carrying reward prediction error signals in the form described by Schultz et al. (1997) for dopamine neurons. However, a midbrain dopaminergic prediction error signal arriving at the striatum might well selectively trigger serotoninergic activity to the extent to which the gamble's prior EV was seen to belong to the loss or punishment domain (while prediction errors in the gain domain would remain dopamine transmitted).

Accordingly, given that the FRN is believed to reflect a dopaminergic signal, we should not be surprised to see it show greater sensitivity to the gain domain. Both Kreussel et al. (2012) and our own experiment showed no significant difference in the FRN for good and bad outcomes in the loss domain. Nevertheless, the bad outcomes in both studies produced more negative FRNs than the good, and other studies (e.g., Crowley, Wu, Bailey, & Mayes, 2009; Holroyd et al., 2004) have suggested the FRN discriminates good and bad outcomes in the loss domain, implying that the differential effect of the domains on the FRN is a matter of degree. This might be due to there being only a partial separation of dopamine and serotonin coding function in gain and loss domains, or because an FRN is generated by serotoninergic activity at the striatum but its more caudal source results in a reduced potential at the scalp.

We also measured the P300, a component believed to be broadly sensitive to the informational content of a stimulus. While it might be expected that the P300 would be greater in the loss domain, as a result of loss aversion, we found that gain domain gambles elicited a larger P300. This accords with Kreussel et al.'s (2012) results (although the P300 fell just short of significance in their study). The P300 is assumed to respond to task-relevant information (e.g., Nieuwenhuis, Aston-Jones, & Cohen, 2005), and so this component may tell us something about how participants perceived the task they were engaged in. If so, participants appear to have perceived gambles in the gain domain as more relevant to the task at hand than those in the loss domain. Note that while net earnings in Experiment 1 were predetermined to be zero, participants' bankrolls increased over time due to the bonus money received from correctly answering probe questions. Once an upward trajectory began to be accepted as the norm, participants may have come to perceive their task as maximizing the speed with which they increased their bankroll, and while loss domain gambles are logically as relevant as gain domain gambles to this task, participants may not have perceived them as such. Kreussel et al. (2012) suggested that this asymmetry of perceived task relevance of the two domains could in fact underlie the FRN sensitivity to gains in their study, which was only observable when (growth over) time was added as an extra variable. In the current experiment, the support for this time-based interpretation of the domain effect was not strong insofar as the three-way Time $\times$ Domain $\times$ Valence interaction was present for only one of the three methods of quantifying the FRN. Furthermore, evidence against Kreussel et al.'s interpretation comes from the SPN. This component is assumed to reflect anticipatory attention (van Boxtel & Bocker, 2004). In our data, the SPN for loss gambles was greater than that for gain gambles, suggesting that participants were more sensitive to prospective losses, in keeping with their later affective ratings. This contrasted with their FRNs, where there was greater sensitivity for gain gambles, as indicated by the greater valence effect in that domain. We therefore do not regard participants' perception of the task as the basis of the domain effect found in the FRN.

The experiments described in this study revealed a discrepancy between the FRN and subsequent affective ratings. Participants' affective ratings were sensitive both to outcome value relative to EV and the absolute value of the outcome. Sensitivity to absolute value was shown by an overall preference for gain domain gambles and for zero-distal gambles, both of which gave the better monetary

return on average. Sensitivity to outcome value relative to EV was shown by good outcomes being preferred to bad. These simultaneous effects are nicely demonstrated by participants' ratings of zero-proximal/good and zero-distal/bad conditions in Experiment 2. The zero-distal/bad outcome was bad compared to its alternative yet returned a higher absolute gain than the zero/proximal good outcome. The sum effect of these dual considerations on participants' happiness appears to be that these two outcomes were rated equally pleasant. At a conscious, reflective level, participants thus appear to rate outcomes using both absolute value and EV. These very strong main effects in both experiments were modulated by significant but weaker interactions consistent with prospect theory.

The FRN was very different. The valence main effect was strong, showing that a comparison with EV was made (reflecting the orthodoxy of Holroyd and Coles' 2002 theory), but the marked main effects of domain and zero proximity that were present in the affective ratings were entirely absent. The FRN does not therefore appear to directly compute absolute outcome value, a finding also shown by Holroyd et al. (2004). Domain did have an effect in the form of an interaction with valence, but this cannot be explained as a direct comparison of the outcome relative to the zero anchor point. First, this should produce a main effect of domain. Furthermore, if this were the case, the observed interaction would have to be driven by one of the outcomes that differed from the zero point, that is, by the gain/good (win 10) or lose/bad (lose 10) outcomes. In fact, when measured by a peak-to-average-peak measure, the Domain × Valence interaction was driven by the gain/bad (win 0) outcome showing a large FRN, while under a mean voltage measure both a large FRN for gain/bad outcomes and a small FRN for gain/good outcomes contributed to the Domain × Valence interaction. As such, the domain effect in the FRN data is best described by an overall increased sensitivity to prediction errors in the gain domain rather than a direct assessment of the outcome relative to zero.

Affective ratings are rarely taken in FRN research; however, Hajcak et al. (2006), Moser and Simons (2009), San Martin et al. (2010), and Yeung and Sanfey (2004) have shown consistency between FRNs and subsequent affective ratings. In contrast, Toyomaki and Murohashi (2005) reported that differences in stated satisfaction with outcomes were not always reflected in participants' FRNs. However, satisfaction was described in purely anecdotal terms in that study. Here, we have quantitatively shown dissociation between a component that putatively measures the hedonic impact of a stimulus at time of delivery and subsequent affective ratings of that stimulus. This may be problematic for claims that the FRN reflects the motivational significance of outcomes (e.g., Gehring & Willoughby, 2002; Luu, Tucker, Derryberry, Reed, & Poulsen, 2003) or constitutes a somatic marker (Holroyd, Nieuwenhuis, Yeung, & Cohen, 2003).

In summary, the generation of anchor point–dependent value judgments in the brain appears to be complex. The existing fMRI literature has not yet resolved where these occur. We present the first ERP experiment that explicitly attempts to study the two anchor point effects described by prospect theory. It is always difficult to integrate ERP and fMRI findings owing to the techniques' differential resolution in the spatial and temporal domains. However, our findings support the view that value coding in the brain is componential, with different structures, and possibly neuromodulators, showing varying sensitivity to the zero anchor point. We reach this conclusion because the behavioral effect of loss aversion, described by prospect theory and behaviorally evident in this experiment, was preceded by an electrophysiological component known to be sensitive to value, which was more sensitive to gains than losses.

References

Bayer, H. M., & Glimcher, P. W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron, 47*, 129–141. doi: 10.1016/j.neuron.2005.05.020

Bayer, H. M., Lau, B., & Glimcher, P. W. (2007). Statistics of midbrain dopamine neuron spike trains in the awake primate. *Journal of Neurophysiology, 98*, 1428–1439. doi: 10.1152/jn.01140.2006

Bellebaum, C., & Daum, I. (2008). Learning-related changes in reward expectancy are reflected in the feedback-related negativity. *European Journal of Neuroscience, 27*, 1823–1835. doi: 10.1111/j.1460-9568.2008.06138.x

Bellebaum, C., Polezzi, D., & Daum, I. (2010). It is less than you expected: The feedback-related negativity reflects violations of reward magnitude expectations. *Neuropsychologia, 48*, 3343–3350. doi: 10.1016/j.neuropsychologia.2010.07.023

Brown, P., & Molliver, M. E. (2000). Dual serotonin (5-HT) projections to the nucleus accumbens core and shell: Relation of the 5-HT transporter to amphetamine-induced neurotoxicity. *Journal of Neuroscience, 20*, 1952–1963.

Brunia, C. H. M., & van Boxtel, G. J. M. (2012). Negative slow waves as indices of anticipation: The bereitschaftspotential, the contingent negative variation, and the stimulus preceding negativity. In S. J. Luck & E. S. Kappenman (Eds.), *Oxford handbook of event-related potential components*. New York, NY: Oxford University Press.

Campbell-Meiklejohn, D., Wakeley, J., Herbert, V., Cook, J., Scollo, P., Ray, M. K., . . . Rogers, R. D. (2011). Serotonin and dopamine play complementary roles in gambling to recover losses. *Neuropsychopharmacology, 36*, 402–410. doi: 10.1038/Npp.2010.170

Carlson, J. M., Foti, D., Mujica-Parodi, L. R., Harmon-Jones, E., & Hajcak, G. (2011). Ventral striatal and medial prefrontal BOLD activation is correlated with reward-related electrocortical activity: A combined ERP and fMRI study. *Neuroimage, 57*, 1608–1616. doi: 10.1016/j.neuroimage.2011.05.037

Chen, M. K., Lakshminarayanan, V., & Santos, L. R. (2006). How basic are behavioral biases? Evidence from capuchin monkey trading behavior. *Journal of Political Economy, 114*, 517–537.

Cohen, M. X., Cavanagh, J. F., & Slagter, H. A. (2011). Event-related potential activity in the basal ganglia differentiates rewards from nonrewards: Temporospatial principal components analysis and source localization of the feedback negativity: Commentary. *Human Brain Mapping, 32*, 2270–2271. doi: 10.1002/Hbm.21358

Crowley, M. J., Wu, J., Bailey, C. A., & Mayes, L. C. (2009). Bringing in the negative reinforcements: The avoidance feedback-related negativity. *Neuroreport, 20*, 1513–1517. doi: 10.1097/Wnr.0b013e32832ff2f5

Daw, N. D., Kakade, S., & Dayan, P. (2002). Opponent interactions between serotonin and dopamine. *Neural Networks, 15*, 603–616.

de Martino, B., Kumaran, D., Holt, B., & Dolan, R. J. (2009). The neurobiology of reference-dependent value computation. *Journal of Neuroscience, 29*, 3833–3842. doi: 10.1523/Jneurosci.4832-08.2009

de Martino, B., Kumaran, D., Seymour, B., & Dolan, R. J. (2006). Frames, biases, and rational decision-making in the human brain. *Science, 313*, 684–687. doi: 10.1126/science.1128356

Foti, D., Weinberg, A., Dien, J., & Hajcak, G. (2011a). Event-related potential activity in the basal ganglia differentiates rewards from nonrewards: Response to Commentary. *Human Brain Mapping, 32*, 2267–2269. doi: 10.1002/Hbm.21357

Foti, D., Weinberg, A., Dien, J., & Hajcak, G. (2011b). Event-related potential activity in the basal ganglia differentiates rewards from nonrewards: Temporospatial principal components analysis and source localization of the feedback negativity. *Human Brain Mapping, 32*, 2207–2216. doi: 10.1002/Hbm.21182

Gehring, W. J., & Willoughby, A. R. (2002). The medial frontal cortex and the rapid processing of monetary gains and losses. *Science, 295*, 2279–2282.

Gu, R. L., Lei, Z. H., Broster, L., Wu, T. T., Jiang, Y., & Luo, Y. J. (2011). Beyond valence and magnitude: A flexible evaluative coding system in the brain. *Neuropsychologia, 49*, 3891–3897. doi: 10.1016/j.neuropsychologia.2011.10.006

Hajcak, G., Moser, J. S., Holroyd, C. B., & Simons, R. F. (2006). The feedback-related negativity reflects the binary evaluation of good versus bad outcomes. *Biological Psychology, 71*, 148–154. doi: 10.1016/j.biopsycho.2005.04.001

Hajcak, G., Moser, J. S., Holroyd, C. B., & Simons, R. F. (2007). It's worse than you thought: The feedback negativity and violations of reward prediction in gambling tasks. *Psychophysiology, 44*, 905–912. doi: 10.1111/j.1469-8986.2007.00567.x

Heidbreder, C. A., Hedou, G., & Feldon, J. (1999). Behavioral neurochemistry reveals a new functional dichotomy in the shell subregion of the nucleus accumbens. *Progress in Neuro-Psychopharmacology & Biological Psychiatry, 23*, 99–132.

Hollerman, J. R., & Schultz, W. (1998). Dopamine neurons report an error in the temporal prediction of reward during learning. *Nature Neuroscience, 1*, 304–309.

Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review, 109*, 679–709. doi: 10.1037//0033-295x.109.4.679

Holroyd, C. B., Larsen, J. T., & Cohen, J. D. (2004). Context dependence of the event-related brain potential associated with reward and punishment. *Psychophysiology, 41*, 245–253. doi: 10.1111/j.1469-8986.2004.00152.x

Holroyd, C. B., Nieuwenhuis, S., Yeung, N., & Cohen, J. D. (2003). Errors in reward prediction are reflected in the event-related brain potential. *Neuroreport, 14*, 2481–2484. doi: 10.1097/01.wnr.0000099601/41403.a5

Kahneman, D., & Tversky, A. (1979). Prospect theory—Analysis of decision under risk. *Econometrica, 47*, 263–291.

Knutson, B., Adams, C. M., Fong, G. W., & Hommer, D. (2001). Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *Journal of Neuroscience, 21*, article RC159.

Knutson, B., Wimmer, G. E., Rick, S., Hollon, N. G., Prelec, D., & Loewenstein, G. (2008). Neural antecedents of the endowment effect. *Neuron, 58*, 814–822. doi: 10.1016/j.neuron.2008.05.018

Kotani, Y., Kishida, S., Hiraku, S., Suda, K., Ishii, M., & Aihara, Y. (2003). Effects of information and reward on stimulus-preceding negativity prior to feedback stimuli. *Psychophysiology, 40*, 818–826.

Kreussel, L., Hewig, J., Kretschmer, N., Hecht, H., Coles, M. G. H., & Miltner, W. H. R. (2012). The influence of the magnitude, probability, and valence of potential wins and losses on the amplitude of the feedback negativity. *Psychophysiology, 49*, 207–219. doi: 10.1111/j.1469-8986.2011.01291.x

Lakshminarayanan, V. R., Chen, M. K., & Santos, L. R. (2011). The evolution of decision-making under risk: Framing effects in monkey risk preferences. *Journal of Experimental Social Psychology, 47*, 689–693. doi: 10.1016/j.jesp.2010.12.011

Liu, X., Hairston, J., Schrier, M., & Fan, J. (2011). Common and distinct networks underlying reward valence and processing stages: A meta-analysis of functional neuroimaging studies. *Neuroscience and Biobehavioral Reviews, 35*, 1219–1236. doi: 10.1016/j.neubiorev.2010.12.012

Luu, P., Tucker, D. M., Derryberry, D., Reed, M. & Poulsen, C. (2003). Electrophysiological responses to errors and feedback in the process of action regulation. *Psychological Science, 14*, 47–53. doi: 10.1111/1467-9280.01417

Marco-Pallares, J., Cucurell, D., Munte, T. F., Strien, N., & Rodriguez-Fornells, A. (2011). On the number of trials needed for a stable feedback-related negativity. *Psychophysiology, 48*, 852–860. doi: 10.1111/j.1469-8986.2010.01152.x

Marco-Pallares, J., Kramer, U. M., Strehl, S., Schroder, A., & Munte, T. F. (2010). When decisions of others matter to me: An electrophysiological analysis. *BMC Neuroscience, 11*, article 86. doi: 10.1186/1471-2202-11-86

Marsh, B., & Kacelnik, A. (2002). Framing effects and risky decisions in starlings. *Proceedings of the National Academy of Sciences of the United States of America, 99*, 3352–3355. doi: 10.1073/pnas.042491999

Martin, L. E., Potts, G. F., Burton, P. C., & Montague, P. R. (2009). Electrophysiological and hemodynamic responses to reward prediction violation. *Neuroreport, 20*, 1140–1143. doi: 10.1097/Wnr.0b013e32832f0dca

Morris, G., Arkadir, D., Nevet, A., Vaadia, E., & Bergman, H. (2004). Coincident but distinct messages of midbrain dopamine and striatal tonically active neurons. *Neuron, 43*, 133–143.

Moser, J. S., & Simons, R. F. (2009). The neural consequences of flip-flopping: The feedback-related negativity and salience of reward prediction. *Psychophysiology, 46*, 313–320. doi: 10.1111/j.1469-8986.2008.00760.x

Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychological Bulletin, 131*, 510–532. doi: 10.1037/0033-2909.131.4.510

Oliveira, F. T. P., McDonald, J. J., & Goodman, D. (2007). Performance monitoring in the anterior cingulate is not all error related: Expectancy deviation and the representation of action-outcome associations. *Journal of Cognitive Neuroscience, 19*, 1994–2004.

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J., & Frith, C. D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature, 442*, 1042–1045. doi: 10.1038/Nature05051

Pine, A., Seymour, B., Roiser, J. P., Bossaerts, P., Friston, K. J., Curran, H. V., & Dolan, R. J. (2009). Encoding of marginal utility across time in the human brain. *Journal of Neuroscience, 29*, 9575–9581. doi: 10.1523/Jneurosci.1126-09.2009

Potts, G. F., Martin, L. E., Burton, P., & Montague, P. R. (2006). When things are better or worse than expected: The medial frontal cortex and the allocation of processing resources. *Journal of Cognitive Neuroscience, 18*, 1112–1119.

San Martin, R., Manes, F., Hurtado, E., Isla, P., & Ibanez, A. (2010). Size and probability of rewards modulate the feedback error-related negativity associated with wins but not losses in a monetarily rewarded gambling task. *Neuroimage, 51*, 1194–1204. doi: 10.1016/j.neuroimage.2010.03.031

Satoh, T., Nakai, S., Sato, T., & Kimura, M. (2003). Correlated coding of motivation and outcome of decision by dopamine neurons. *Journal of Neuroscience, 23*, 9913–9923.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275*, 1593–1599.

Seymour, B., Daw, N., Dayan, P., Singer, T., & Dolan, R. (2007). Differential encoding of losses and gains in the human striatum. *Journal of Neuroscience, 27*, 4826–4831. doi: 10.1523/Jneurosci.0400-07.2007

Sutton, R., & Barto, A. (1998). *Reinforcement learning: An introduction.* Cambridge, MA: MIT Press.

Tom, S. M., Fox, C. R., Trepel, C., & Poldrack, R. A. (2007). The neural basis of loss aversion in decision-making under risk. *Science, 315*, 515–518. doi: 10.1126/science.1134239

Toyomaki, A., & Murohashi, H. (2005). Discrepancy between feedback negativity and subjective evaluation in gambling. *Neuroreport, 16*, 1865–1868.

Trepel, C., Fox, C. R., & Poldrack, R. A. (2005). Prospect theory on the brain? Toward a cognitive neuroscience of decision under risk. *Cognitive Brain Research, 23*, 34–50. doi: 10.1016/j.cogbrainres.2005.01.016

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

Tversky, A., & Kahneman, D. (1992). Advances in prospect-theory—Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5*, 297–323.

van Boxtel, G. J. M., & Bocker, K. B. E. (2004). Cortical measures of anticipation. *Journal of Psychophysiology, 18*, 61–76. doi: 10.1027/0269-8803.18.2-3.61

von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.

Weber, B., Aholt, A., Neuhaus, C., Trautner, P., Elger, C. E., & Teichert, T. (2007). Neural evidence for reference-dependence in real-market-transactions. *Neuroimage, 35*, 441–447. doi: 10.1016/j.neuroimage.2006.11.034

Wu, Y., & Zhou, X. L. (2009). The P300 and reward valence, magnitude, and expectancy in outcome evaluation. *Brain Research, 1286*, 114–122. doi: 10.1016/j.brainres.2009.06.032

Yacubian, J., Glascher, J., Schroeder, K., Sommer, T., Braus, D. F., & Buchel, C. (2006). Dissociable systems for gain- and loss-related value predictions and errors of prediction in the human brain. *Journal of Neuroscience, 26*, 9530–9537. doi: 10.1523/jneurosci.2915-06.2006

Yeung, N., Holroyd, C. B., & Cohen, J. D. (2005). ERP correlates of feedback and reward processing in the presence and absence of response choice. *Cerebral Cortex, 15*, 535–544. doi: 10.1093/cercor/bhh153

Yeung, N., & Sanfey, A. G. (2004). Independent coding of reward magnitude and valence in the human brain. *Journal of Neuroscience, 24*, 6258–6264. doi: 10.1523/Jneurosci.4537-03.2004

Zhong, S. F., Israel, S., Xue, H., Sham, P. C., Ebstein, R. P., & Chew, S. H. (2009). A neurochemical approach to valuation sensitivity over gains and losses. *Proceedings of the Royal Society B: Biological Sciences, 276*, 4181–4188. doi: 10.1098/rspb.2009.1312

(Received April 27, 2012; Accepted August 28, 2012)