

# Available alternative incentives modulate anticipatory nucleus accumbens activation

Jeffrey C. Cooper, Nick G. Hollon, G. Elliott Wimmer, and Brian Knutson

Department of Psychology, Stanford University, Jordan Hall, 450 Serra Mall, Stanford, CA 94305, USA

**A reward or punishment can seem better or worse depending on what else might have happened. Little is known, however, about how neural representations of an anticipated incentive might be influenced by the available alternatives. We used event-related fMRI to investigate the activation in the nucleus accumbens (NAcc), while we varied the available alternative incentives in a monetary incentive delay task. Some task blocks included only uncertain gains and losses; others included the same uncertain gains and losses intermixed with certain gains and losses. The availability of certain gains and losses increased NAcc activation for uncertain losses and decreased the difference between uncertain gains and losses. We suggest that this pattern of activation can result from reference point changes across blocks, and that the worst available loss may serve as an important anchor for NAcc activation. These findings imply that NAcc activation represents anticipated incentive value relative to the current context of available alternative gains and losses.**

**Keywords:** context; reference points; reward; nucleus accumbens; fMRI

## INTRODUCTION

The subjective value of a gain or loss can depend on the availability of other incentives. These changes in value are reflected both in behavioral preferences (Luce, 1992; Simonson and Tversky, 1992) and in neural regions where activation is sensitive to subjective value (Knutson *et al.*, 2001a; Montague *et al.*, 2006; O'Doherty, 2004; Schultz, 2006). For example, human functional magnetic resonance imaging (fMRI) studies have found that value-sensitive areas such as the medial prefrontal cortex, amygdala, striatum and posterior cingulate are more activated by the same monetary outcome (e.g. 50 pence or \$0) when it is the best possible outcome predicted by a cue than when it is the worst (Knutson *et al.*, 2003; Nieuwenhuis *et al.*, 2005; Elliott *et al.*, 2008).

These studies suggest that neural representations of value are relative to at least one other possible outcome on that trial. However, three specific questions have not been addressed by previous research. First, earlier studies of relative value have focused on alternatives within a trial (i.e. was this outcome the best or worst that could have happened in this trial?). Alternatives, though, could also be viewed in a larger context, such as a block of trials (i.e. was this incentive the best or worst incentive that could have been presented in this block?).

Received 6 March 2009; Accepted 16 July 2009

Advance Access publication 20 October 2009

We thank Gregory R. Samanez-Larkin, Kacey Ballard and the rest of the SPAN lab for assistance at all stages of this study and many helpful comments and discussions; Antonio Rangel for many insights; and the anonymous reviewers whose comments greatly improved this study. This work was supported by the Stanford School of Humanities & Sciences Vice Provost of Graduate Education [Graduate Research Opportunity grant to J.C.C.]; the National Institute of Mental Health [5T32MH020006-10 to J.C.C.] and the National Science Foundation [0748915 to B.K.].

Correspondence should be addressed to Jeffrey C. Cooper, Department of Psychology, Jordan Hall, 450 Serra Mall, Stanford, CA 94305, USA. Email: jcooper@stanford.edu.

Little is known about whether or how this broader context might influence neural representations of value. One possibility is a 'range' hypothesis that suggests neural representations of value expand or contract in range so that the best available incentive (regardless of its actual magnitude) elicits maximal activation, while the worst elicits minimal activation. According to this account, changing the value of the best and worst alternatives should expand or contract the range of activation for all alternative incentives. Another possibility is an 'anchor' hypothesis that neural representations of a value are relative to either the best or the worst possible incentive. According to this account, changing the value of the best and worst alternatives should shift activation for all alternative incentives in a single direction. This account does not necessarily specify which of the best or worst would serve as an anchor, but one possibility is that a large potential loss might be a more attention-getting anchor than a gain of equal size (Tversky and Kahneman, 1992).

Second, earlier studies of relative value have only focused on neural responses to neutral outcomes or monetary gains, rather than losses. Many studies suggest that value-sensitive brain areas represent losses differently than gains (Breiter *et al.*, 2001; Bayer and Glimcher, 2005; Seymour *et al.*, 2007; Tom *et al.*, 2007). Changes in alternative available incentives might thus affect the representations of anticipated losses differently than gains. For example, many studies using monetary incentive delay (MID) tasks have demonstrated greater anticipatory nucleus accumbens (NAcc) activation for gains than for losses (Knutson *et al.*, 2001a; Bjork *et al.*, 2004; Knutson *et al.*, 2005; Guyer *et al.*, 2006). A recent study, however, found that when uncertain gains and losses were presented in blocks of trials that also included certain gains and losses, the NAcc was not

differentially activated for uncertain gains *vs* losses, suggesting that activation for losses was sensitive to the available certain incentives (Cooper and Knutson, 2008). This study, though, did not manipulate available alternative incentives within individuals.

Third, earlier studies of relative value have focused on neural responses to receiving outcomes within a trial. Different brain areas, however, may represent value before compared with after an incentive is received (Knutson *et al.*, 2001b; Knutson and Greer, 2008). For example, the NAcc is more active during reward anticipation than in response to reward outcomes, and plays a key role in promoting approach behavior (Ikemoto and Panksepp, 1999; Breiter *et al.*, 2001; Knutson and Cooper, 2005). The influence of changing available incentives on representations of value during anticipation (rather than in response to outcomes) is still unexplored.

To address the questions of whether changing available alternative incentives could change anticipatory neural responses within individuals, and of whether that change would differ between gains and losses, we scanned participants as they performed a modified MID task. Participants faced two sets of available incentives in alternating blocks. In one set, uncertain \$5.00 gains and losses were the best and worst available outcomes, respectively. In the other set, identical uncertain \$5.00 gains and losses were mixed with certain \$5.00 gains and losses, respectively. Using certain gains and losses as the alternative incentives provided a link to studies that have examined neural representations of subjective value using gambles that combine certain and uncertain outcomes (Preuschoff *et al.*, 2006; Bjork and Hommer, 2007; Seymour *et al.*, 2007). More broadly, whether the availability of certain outcomes affects the representations of uncertain outcomes is important for understanding how the brain encodes value in mixed gambles, a commonly used choice format in economics, psychology and neuroscience.

Because of our focus on anticipation, analyses targeted the NAcc. We compared anticipatory NAcc activation to gain and loss cues between blocks with different available alternative incentives. The range hypothesis predicted that when better (certain \$5.00 gains) and worse (certain \$5.00 losses) incentives were available, activation for uncertain gains should decrease, and activation for uncertain losses should increase. In contrast, the anchor hypothesis predicted that when better and worse incentives were available, activation for uncertain gains and losses should both increase (if the worst incentive was the anchor) or both decrease (if the best incentive was the anchor). Based on earlier findings (Cooper and Knutson, 2008) and findings based on loss aversion, we predicted that the worst available incentive would serve as the anchor, and hence that anticipatory NAcc activation for both uncertain gains and uncertain losses would increase when better and worse incentives were available.

## MATERIALS AND METHODS

### Participants

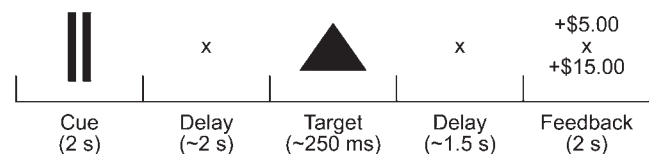
Twelve right-handed healthy volunteers, with no history of neurological or psychiatric diagnosis, participated (six women, aged 18–28 years). Participants gave informed consent for a protocol following the Declaration of Helsinki, and the study was approved by the Institutional Review Board of the Stanford University School of Medicine. Two participants were given incomplete post-experimental questionnaires, and so retrospective affect ratings were only available for 10 participants.

### Experimental design and task

We combined two MID task versions from Knutson *et al.* (2001a) and Cooper and Knutson (2008; Figure 1). In each trial of the task, participants first saw a shape (the ‘cue’; 2000 ms) that indicated the experimental condition. The cue then disappeared during a delay (randomized between 1500 and 2500 ms). A triangle target then appeared briefly (variable between 150 and 470 ms), and participants attempted to ‘hit’ the target with a button press while it was on-screen. After another delay (variable between 1030 and 2350 ms), participants saw the amount they earned on that trial and the total amount they had earned in that block (the ‘outcome’; 2000 ms). Trial length was 8000 ms.

Target speed was adjusted during the experiment by an adaptive timing algorithm that attempted to maintain a constant hit rate for each condition. This algorithm sped up or slowed down the target for each condition if that condition’s hit rate exceeded or fell <66%. All conditions began with 250-ms target duration. Mean target duration over all participants and conditions was 235.7 ms (s.e.m. 4.3 ms).

Participants faced the two incentive sets in separate blocks in the experiment, each with unique cues. The ‘uncertain-only’ (UNC-ONLY) set mimicked the most common version of the MID task (Knutson *et al.*, 2001a). This set contained four conditions, all with uncertain outcomes, that crossed two levels of magnitude (high or low) with two types of incentive valence (gain or loss). Each condition was signaled by a unique cue (e.g. two vertical lines). On high-gain trials, hitting the target earned \$5.00, but misses earned \$0.00 (i.e. no change). On low-gain trials, hits earned \$0.05, but misses earned \$0.00. On high-loss trials, hits earned



**Fig. 1** Trial structure and available incentives. Conditions were distinguished by different cues. During UNC-ONLY blocks, four cues were used; during CERT-UNC blocks, eight different cues were used. Outcome amounts were given both for the current trial and the total earned within the current block.

\$0.00 (i.e. they avoided a loss), but misses lost \$5.00. On low-loss trials, hits earned \$0.00, but misses lost \$0.05.

The 'certain and uncertain' (CERT-UNC) set mimicked the modified version of the MID task used in Cooper and Knutson (2008). This set had eight conditions that crossed two levels of magnitude (high or low) with two types of incentive valence (gain or loss) and two levels of certainty (uncertain or certain). Each condition was signaled by a unique cue (e.g. an empty black square, a white circle within a circle). The four uncertain conditions (uncertain high-gain, uncertain low-gain, uncertain high-loss and uncertain low-loss) were identical to the conditions in the UNC-ONLY set; for example, on uncertain high-gain trials, hitting the target earned \$5.00, but misses earned \$0.00. In the four certain conditions, though, both hits and misses earned the same amount. On certain high-gain trials, both hits and misses earned \$5.00. On certain low-gain trials, both hits and misses earned \$0.05. On certain high-loss trials, both hits and misses lost \$5.00, and on certain low-loss trials, both hits and misses lost \$0.05.

Thus, in the UNC-ONLY set, the best cue signaled an uncertain chance of gaining \$5.00, while the worst cue signaled an uncertain chance of losing \$5.00. In the CERT-UNC set, participants faced cues for identical incentives as the UNC-ONLY set, intermixed with an even better gain cue (signaling a certain \$5.00 gain) and an even worse loss cue (signaling a certain \$5.00 loss). The cues for uncertain \$5.00 gains therefore had identical absolute value across sets but higher relative value in the UNC-ONLY set (when they were the best possible cue), while cues for uncertain \$5.00 losses had identical absolute value across sets but lower relative value in the UNC-ONLY set (when they were the worst possible cue).

Participants played two blocks of each set in the experiment for four total blocks. Sets alternated between each block, with the initial set counterbalanced across participants (e.g. CERT-UNC/UNC-ONLY/CERT-UNC/UNC-ONLY). Each block was scanned in a separate scanner run. Participants were informed as to which set they would face at the beginning of each block. All UNC-ONLY and CERT-UNC blocks had eight trials per condition, for a total of 16 trials per condition. Because there were twice as many CERT-UNC conditions as UNC-ONLY conditions, CERT-UNC blocks took 8.5 min and UNC-ONLY blocks took 4.3 min. Conditions within a block were presented in pseudorandom order individualized for each participant.

Before scanning, participants received instruction on both sets and the meaning of each cue, followed by a 10-min training version of the task including both sets and all conditions. Participants were then shown the cash they could win. Participants were told that they would play two blocks of each set, but were not told how many trials they would face or about the adaptive timing algorithm.

After the experiment, participants rated how they felt during the experiment about each cue on two affective

dimensions, arousal and positivity (Russell, 1980). Each rating was made on 7-point Likert scale; the arousal scale ranged from not aroused to very aroused, while the positivity scale ranged from negative to neutral to positive. Scores were converted such that higher scores represented increasing arousal and increasing positivity. Participants were then paid their total winnings in cash from either the CERT-UNC or the UNC-ONLY set, determined randomly after the experiment. These winnings were added to a participation fee (that ranged from \$45 to \$53, depending on the length of the scan). Total payment thus ranged from \$51 to \$88 (mean \$73.58, s.e.m. \$3.09). Care was taken to ensure that participants regarded the two sets as separate. In addition to paying the winnings for only one set, all instructions referred to the sets as two different games and the cues for each set were distinct from the other set's cues.

Behavioral data were analyzed with SPSS 14.0 and Microsoft Excel 2003 for Windows. In earlier studies, anticipatory NAcc activation for cues correlated with a combination of arousal and positivity ratings for the cue, or 'positive arousal' (Knutson *et al.*, 2001a; Bjork *et al.*, 2004; Knutson and Peterson, 2005). We therefore calculated positive arousal (PA) scores for each condition within each participant. PA scores were calculated for each condition by subtracting the participant's mean arousal or positivity over conditions from that condition's arousal or positivity, summing the mean-deviated arousal and positivity scores, and multiplying by  $\sin(45^\circ)$ . This transformation represents a diagonal rotation of the arousal and positivity axes to a single PA axis (Watson *et al.*, 1999). Reaction times were transformed with the natural logarithm function before analysis to account for their skew. Differences for behavioral measures were tested with repeated-measures ANOVA over both sets using condition as a single within-participant factor (using the Huynh-Feldt correction for non-sphericity, denoted as  $P_{H-F}$ ). The t-tests were used to investigate significant effects *post hoc*.

### Imaging

Participants were scanned with a General Electric 1.5 T Signa scanner using the standard head coil. Stimuli were presented with E-Prime 1.1 and projected on a mirror mounted on the coil. Participants were fitted with a bite bar and padding to minimize head motion. Functional images covered the whole brain and consisted of 24 contiguous 4-mm thick axial slices ( $TR = 2000$  ms,  $TE = 40$  ms,  $\text{flip} = 90^\circ$ ,  $3.44 \times 3.44$  mm in-plane resolution,  $64 \times 64$  matrix), collected using a T2\*-sensitive spiral in/out pulse sequence that minimizes dropout in ventral frontal and medial temporal regions (Glover and Law, 2001; Preston *et al.*, 2004). One hundred and thirty-three images were collected in each UNC-ONLY run, while 261 images were collected in each CERT-UNC run, for a total of 788 images; the first five images of each run were discarded to allow for

magnetic equilibration. An in-plane structural image was acquired before the functional runs (24 contiguous 4-mm thick axial slices; TR = 14 ms; TR = 400 ms,  $0.94 \times 0.94$  mm in-plane resolution,  $256 \times 256$  matrix), and a high-resolution structural was acquired after the functional runs (3D acquisition; T1-weighted SPGR sequence;  $0.86 \times 0.86 \times 1.5$  mm voxel size;  $256 \times 256 \times 116$  matrix).

### Region of interest data analysis

The specific hypotheses for this study concerned NAcc activation, and so they were tested using a region of interest (ROI) approach. ROI masks for the NAcc were based on each participant's individual anatomy (Supplementary Figure S1; Breiter *et al.*, 1997); this method avoids biasing the ROI specification with the functional data (Devlin and Poldrack, 2007; Kriegeskorte *et al.*, 2009). Left and right NAcc ROIs were located on each participant's high-resolution structural scan by placing 8-mm diameter spheres at a starting point of  $x/y/z = \pm 12/10/-2$  mm and shifting them to ensure each sphere sampled only gray matter. No sphere was adjusted for  $>7$  mm in any direction. Marsbar was used to process ROI data (Brett *et al.*, 2002). Functional data were extracted, averaged across all voxels in each ROI and filtered with a high-pass filter (cut-off 90 s). The time course of responses to each cue in each ROI was then modeled with a finite impulse response model (averaging over both blocks of each incentive set) and estimated using restricted maximum likelihood, correcting for temporal autocorrelation with an AR(1) model. Time courses were converted to percent signal change from the experiment mean in each ROI and averaged across participants. We first analyzed left and right NAcc time courses with repeated measures ANOVA including hemisphere (left or right), condition (12 levels) and time point (six levels from 0 to 10 s) as factors. Hemisphere showed no main effect or interactions (all  $F$ 's  $< 1$ ), so we combined left and right time courses into a single bilateral NAcc time course for all further analyses. Because our hypotheses specifically concerned anticipatory activation in response to the cue, repeated measures ANOVA and planned  $t$ -tests at the expected hemodynamic response peak of 4 s after cue onset were used to investigate significant differences between conditions (see supplementary results for similar tests at 6 s and of parameter estimates).

### Whole-brain data analysis

Whole-brain imaging data were preprocessed and modeled with SPM2 (Wellcome Department of Imaging Neuroscience). Functional images were corrected for slice timing and realigned to the first image in the run. In-plane and high-resolution scans were co-registered to the mean functional image and normalized to the MNI avg152 template brain using standard parameters (a first pass of 12-parameter affine normalization, followed by non-linear transformation with discrete cosine transform basis

functions, using a 25-mm cut-off and 16 non-linear iterations). Functional images were normalized with the in-plane parameters, interpolated to  $2 \times 2 \times 2$ -mm voxels, and smoothed with a 4-mm FWHM Gaussian filter. A high-pass filter (cut-off 90 s) was applied within runs to remove low-frequency noise.

Experimental effects were modeled within each participant using a general linear model. Each model included regressors for each condition's cue and outcome events (with 1 at event onsets and 0 otherwise). We also included a first-order parametric modulator at each cue weighted with reaction time on that trial to account for trial-to-trial variation in motor response (Knutson *et al.*, 2005). All regressors of interest and the reaction time modulator were convolved with a single gamma function model of the hemodynamic response (Cox, 1996). Six regressors modeling residual head motion ( $x$ ,  $y$ ,  $z$  and pitch, roll, yaw) and a constant term within each run were also included. Restricted maximum likelihood estimation was used to create whole-brain  $\beta$ -weight images, correcting for temporal autocorrelation with an AR(1) model. The  $\beta$ -weight images were combined with appropriate weights to form contrast images for each participant. Group effects for each contrast were then tested with one-sample  $t$ -tests over all participants' contrast images. Significant voxels were identified with a voxelwise threshold of  $P < 0.001$  and a cluster size threshold of greater than 16 voxels ( $128 \text{ mm}^3$ ), providing whole-brain protection against false positives at  $P < 0.05$  according to AlphaSim (Ward, 2002). Peak activations are reported in MNI coordinates, as in SPM2.

## RESULTS

### Behavior

Participants responded on almost all trials (mean response rate = 93.7%, s.e.m. = 0.01%) and there were no significant differences between conditions in response rate [ $F(11, 121) = 0.70$ ,  $P > 0.73$ ]. Hit rates also did not significantly differ between conditions [mean hit rate = 65.9%, s.e.m. = 0.01%,  $F(11, 121) = 0.18$ ,  $P > 0.99$ ], confirming that the adaptive timing algorithm controlled hit rates. Mean reaction time was 199.99 ms (s.e.m. = 3.88 ms), but reaction times did significantly differ across conditions [ $F(11, 121) = 2.50$ ,  $P_{H-F} < 0.05$ ]. The  $t$ -tests indicated that these differences were largely due to relatively faster responses for the uncertain high-magnitude conditions (Table 1). No uncertain condition's reaction time differed between contexts.

Retrospective affect ratings also differed by condition [arousal:  $F(11, 99) = 9.30$ ,  $P_{H-F} < 0.001$ ; positivity:  $F(11, 99) = 14.70$ ,  $P_{H-F} < 0.001$ ]. For arousal,  $t$ -tests indicated that these differences were due to higher arousal for high-magnitude than low-magnitude trials, especially for uncertain high-magnitude trials (Table 2). For positivity, differences were driven by greater positivity for gains than losses, especially for high-magnitude trials. No uncertain

**Table 1** Reaction time by condition

Condition	Reaction time (s.e.m.)
UNC-ONLY	
Uncertain high-gain	192.5 <sub>a,b,c</sub> (3.7)
Uncertain low-gain	198.7 <sub>a,c,d,e,f</sub> (4.9)
Uncertain low-loss	197.9 <sub>a,b,c,d,e</sub> (7.0)
Uncertain high-loss	194.0 <sub>b,g</sub> (4.8)
CERT-UNC	
Uncertain high-gain	193 <sub>a,b,c,e</sub> (4.6)
Uncertain low-gain	202.2 <sub>c,f,g</sub> (5.7)
Uncertain low-loss	208 <sub>a,b,c,f</sub> (7.3)
Uncertain high-loss	192.8 <sub>a,b</sub> (3.7)
Certain high-gain	204.8 <sub>d,f</sub> (5.3)
Certain low-gain	203.3 <sub>a,b,c,d,e</sub> (8.3)
Certain low-loss	207 <sub>f</sub> (4.4)
Certain high-loss	205.6 <sub>e,f</sub> (4.4)

Note: Data are mean reaction time in milliseconds. Reaction times were log-transformed for statistical comparison; original reaction times are reported here for clarity. The s.e.m. are calculated within condition. Data points that share subscripts do not differ at  $P < 0.05$  (two tailed).

**Table 2** Cue affect ratings by condition

Condition	Arousal (s.e.m.)	Positivity (s.e.m.)	PA (s.e.m.)
UNC-ONLY			
Uncertain high-gain	6.4 <sub>a</sub> (0.2)	5.8 <sub>ij</sub> (0.2)	2.1 <sub>r</sub> (0.2)
Uncertain low-gain	4.1 <sub>c,d,e,f</sub> (0.5)	4.5 <sub>kl</sub> (0.4)	-0.4 <sub>u,v,z</sub> (0.4)
Uncertain low-loss	4.0 <sub>c,d,e,f</sub> (0.4)	3.5 <sub>m,n</sub> (0.3)	-1.2 <sub>x,z</sub> (0.2)
Uncertain high-loss	6.1 <sub>a</sub> (0.3)	2.5 <sub>o,p</sub> (0.3)	-0.4 <sub>t,v,x,y</sub> (0.3)
CERT-UNC			
Uncertain high-gain	5.4 <sub>a,b,c</sub> (0.5)	5.3 <sub>jk</sub> (0.3)	1.6 <sub>r,s</sub> (0.5)
Uncertain low-gain	3.8 <sub>d,f</sub> (0.4)	4.3 <sub>km</sub> (0.3)	-0.3 <sub>u,w,y</sub> (0.3)
Uncertain low-loss	4.3 <sub>b,e</sub> (0.4)	3.3 <sub>l,m,o</sub> (0.4)	-0.6 <sub>w,x,z</sub> (0.3)
Uncertain high-loss	5.9 <sub>a</sub> (0.3)	2.8 <sub>n,o,p</sub> (0.4)	0.1 <sub>s,u,w</sub> (0.4)
Certain high-gain	4.9 <sub>b,d</sub> (0.5)	6.0 <sub>i</sub> (0.3)	1.7 <sub>r</sub> (0.4)
Certain low-gain	3.9 <sub>c,e,f</sub> (0.5)	4.6 <sub>k</sub> (0.3)	0.0 <sub>t,w</sub> (0.3)
Certain low-loss	3.7 <sub>e,f</sub> (0.4)	3.0 <sub>n,o,p</sub> (0.3)	-1.3 <sub>z</sub> (0.2)
Certain high-loss	4.5 <sub>b,c</sub> (0.4)	2.3 <sub>p</sub> (0.4)	-1.2 <sub>y,z</sub> (0.3)

Note: Data are mean retrospective affect ratings ( $n = 10$ ) of each condition's cue on 7-point Likert scales for arousal and positivity. PA was calculated from arousal and positivity (see text for formula). The s.e.m. are calculated within condition. Data points within a column that share subscripts do not differ at  $P < 0.05$  (two tailed).

conditions differed between UNC-ONLY and CERT-UNC sets in either arousal or positivity. PA scores also differed by trial condition [ $F(11, 99) = 11.76$ ,  $P_{H-F} < 0.001$ ];  $t$ -tests indicated these differences were due to greater PA in high-gain trials (both uncertain and certain) than in high-loss or low-magnitude trials. Uncertain high-loss trials also had slightly higher PA in the UNC-ONLY set than in the CERT-UNC set [ $t(9) = 2.40$ ,  $P < 0.05$ ].

### ROI analyses

In order to test our hypotheses about anticipatory NAcc activation, we estimated average percent signal change time

courses in the NAcc for all conditions. We then analyzed activity in the uncertain trials (which were present in both contexts) at 4 s following the cue for effects of available incentive set (UNC-ONLY or CERT-UNC), valence (gain or loss) and magnitude (high or low; Figure 2 and Table 3). There were significant main effects of set [ $F(1, 11) = 10.01$ ,  $P_{H-F} < 0.01$ ] and magnitude [ $F(1, 11) = 7.29$ ,  $P_{H-F} < 0.05$ ], a significant interaction between valence and magnitude [ $F(1, 11) = 5.31$ ,  $P_{H-F} < 0.05$ ] and a significant three-way interaction of set, valence and magnitude [ $F(1, 11) = 10.50$ ,  $P_{H-F} < 0.01$ ].

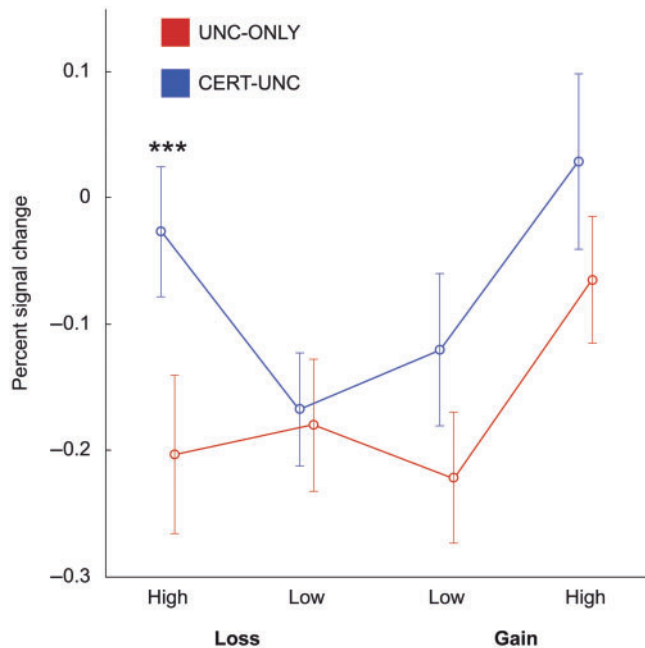
We investigated the effect of available incentive set with planned  $t$ -tests between uncertain high-magnitude gains and losses in each set. These trials had identical absolute values but differing relative values across sets. In the CERT-UNC set compared with the UNC-ONLY set, uncertain gains had lower relative value, while uncertain losses had higher relative value. Uncertain high-loss trials had significantly greater activation in the CERT-UNC than in the UNC-ONLY set [ $t(11) = 4.89$ ,  $P < 0.001$ ]. Uncertain high-gain trials had a trend toward greater activation in the CERT-UNC set than in the UNC-ONLY set [ $t(11) = 1.96$ ,  $P = 0.076$ , two tailed]. Among other uncertain trials, no low-magnitude trial differed from any other in either set, and they had lower activation than most high-magnitude trials.

The valence by magnitude interaction was evident primarily in the UNC-ONLY set, where uncertain high-gain trials elicited greater activation than uncertain high-loss trials [ $t(11) = 3.53$ ,  $P < 0.01$ ]. This difference was not present in the CERT-UNC set [ $t(11) = 1.25$ ,  $P > 0.23$ ], which accounted for the three-way interaction.

In summary, anticipatory NAcc activation for a given incentive was significantly influenced by which alternative incentives were available. In the CERT-UNC set, activation was greater for uncertain losses, and the difference between gains and losses was significantly reduced.

As this study aimed to investigate the effects of incentive set, we did not focus the analyses on certain trials, which appeared only in the CERT-UNC set. Exploratory analyses revealed a few differences from other conditions (Table 3). In particular, NAcc activation on certain high-gain trials did not differ from any other condition. NAcc activation on certain high-loss trials, however, was lower than uncertain high-gain trials in both the UNC-ONLY [ $t(11) = 3.09$ ,  $P < 0.05$ ] and CERT-UNC [ $t(11) = 4.21$ ,  $P < 0.01$ ] sets, as well as lower than uncertain high-loss trials in the CERT-UNC set [ $t(11) = 3.55$ ,  $P < 0.01$ ] and certain low-gain trials [ $t(11) = 3.19$ ,  $P < 0.01$ ].

Similar results were found for activation at 6 s following cue onset and parameter estimates of activation in response to cues (Supplementary Results).



**Fig. 2** Percent signal change in NAcc. Points represent mean percent signal change from the experiment mean in bilateral NAcc at 4 s following cue onset. Error bars represent standard errors within participants. Only uncertain trials are shown, and only between-set differences for the same incentive are indicated. See Table 3 for data and all pairwise comparisons. \*\*\* $p < 0.001$ .

**Table 3** Peak NAcc activation by condition

Condition	Percent signal change (s.e.m.)
UNC-ONLY	
Uncertain high-gain	-0.065 <sub>a,b,c</sub> (0.05)
Uncertain low-gain	-0.22 <sub>e</sub> (0.05)
Uncertain low-loss	-0.18 <sub>c,d,e</sub> (0.05)
Uncertain high-loss	-0.20 <sub>d,e</sub> (0.06)
CERT-UNC	
Uncertain high-gain	0.029 <sub>a</sub> (0.07)
Uncertain low-gain	-0.12 <sub>b,c,d,e</sub> (0.06)
Uncertain low-loss	-0.17 <sub>d,e</sub> (0.05)
Uncertain high-loss	-0.027 <sub>a,b</sub> (0.05)
Certain high-gain	-0.11 <sub>a,b,c,d,e</sub> (0.05)
Certain low-gain	-0.091 <sub>b,c,d</sub> (0.05)
Certain low-loss	-0.16 <sub>b,c,d,e</sub> (0.04)
Certain high-loss	-0.19 <sub>e</sub> (0.04)

Note: Data are activation in the NAcc 4 s following the cue for a condition, in units of percent signal change from the experiment mean in that ROI. The s.e.m. are calculated within condition. Data points that share subscripts do not differ at  $P < 0.05$  (two tailed).

### Whole-brain analyses

Although this study was designed to specifically focus on NAcc activation, we also performed whole-brain analyses of anticipatory activation to cue onset between conditions (see Supplementary Tables S3–S5 for full activation list and Supplementary Figure S3 for image). In both the UNC-ONLY and CERT-UNC sets, uncertain high-gain trials

compared with uncertain low-gain trials activated a wide network of medial and anterior frontal, striatal and parietal areas commonly associated with motivation and valuation (Knutson and Cooper, 2005; Montague *et al.*, 2006; Schultz, 2006). As well, in both sets, uncertain high-loss trials compared with uncertain low-loss trials activated a smaller but partially overlapping network of brain areas, including dorsal and posterior cingulate and dorsal striatal areas (Seymour *et al.*, 2007).

Across incentive sets, uncertain high-gain trials elicited more activation in the UNC-ONLY set compared with the CERT-UNC set in anterior prefrontal cortex and in the temporal pole. In the CERT-UNC compared with the UNC-ONLY set, uncertain high-gain trials elicited more activation in inferior frontal gyrus. No clusters were differentially active between incentive sets for uncertain high-loss trials.

### DISCUSSION

The current findings demonstrate that in the NAcc, a crucial human substrate for representing anticipated reward, anticipatory activation for an incentive is modulated by the available alternative incentives. Using event-related fMRI and a version of the MID task (Knutson *et al.*, 2001a) incorporating two different sets of incentives, we compared NAcc activation to cues for uncertain \$5 gains and losses when their relative (but not absolute) value differed. Specifically, uncertain gains and losses were the best and worst incentives in one set (UNC-ONLY), but the second-best and second-worst in another set (CERT-UNC).

For the first time, we observed that NAcc activation differed during anticipation of losses (and to a weaker extent gains) that were identical in absolute magnitude, probability, expected value and uncertainty. The losses and gains differed only in which other incentives were available during that block of trials. When uncertain gains and losses were the best and worst incentives, NAcc activation was significantly greater for uncertain gains than for uncertain losses. However, when the same uncertain gains and losses were not the best and worst incentives, NAcc activation for uncertain losses increased, and the difference between activation for uncertain gains and losses was reduced. This effect was present by 4 s following cue onset, but similar patterns for activation at 6 s and for parameter estimates of the hemodynamic response function indicated the effect extended throughout the anticipation period.

The set of available alternative incentives, then, modulated a key neural representation of anticipated incentive value in accordance with the anchor hypothesis as opposed to the range hypothesis. Activation for uncertain losses increased when they were not the worst outcome, while activation for uncertain gains also showed a trend toward an increase. Since choice for consumer products can be predicted with NAcc activation (Knutson *et al.*, 2007), these findings imply that preference reversals across different incentive sets

(Tversky and Simonson, 1993) may relate to changes in the NAcc representation of anticipated value.

Although this study focused on NAcc activation, we also conducted whole-brain analyses both within and across incentive sets. Within each set, we found that common networks involved in motivation and value were recruited for high-gain and high-loss trials, including medial frontal, striatal and parietal areas. Across sets, a few clusters were differently active for uncertain high-gain trials (but none for uncertain high-loss trials): right anterior PFC and temporal pole were more active in the UNC-ONLY set, while inferior frontal gyrus was more active in the CERT-UNC set. Importantly, the small set of brain areas differentially active between incentive sets suggest that differences in the NAcc cannot be explained by an overall increase in activation across the whole brain.

These findings go beyond existing studies of context on activation in the NAcc, to directly test competing hypotheses of how relative value is represented. Earlier studies (Knutson *et al.*, 2003; Nieuwenhuis *et al.*, 2005; see also Tobler *et al.*, 2005; Elliott *et al.*, 2008) suggested that value-sensitive areas may represent the value of an outcome ranked relative to other available outcomes. In this study, both better and worse incentives were added in the CERT-UNC set. The range hypothesis suggested that activation for uncertain gains would decrease and activation for uncertain losses would increase in this set, as they could now be judged relative to both better (certain gain) and worse (certain loss) alternatives. The anchor hypothesis, though, suggested that uncertain gains and losses would both be judged relative to only one of the new incentives, and hence that activation for both uncertain gains and losses would increase or decrease.

These findings support the anchor hypothesis (specifically for an anchor at the worst possible incentive). Activation for uncertain losses, and to a lesser extent uncertain gains, increased after adding a worse and a better incentive to a block of trials. This suggests that in the CERT-UNC set, anticipatory NAcc activation was influenced more by the availability of a certain loss than by a certain gain of the same magnitude, and hence activation for both uncertain incentives was increased. The fact that the worst available incentive, as opposed to the best, seemed to serve as the anchor is consistent with the loss-aversion observation that losses can have a greater impact than gains of the same magnitude.

An additional possibility is that NAcc representations of uncertain losses might be more affected by other available incentives because the NAcc represents anticipated gains differently from anticipated losses (Breiter *et al.*, 2001; Knutson *et al.*, 2001a). This account is supported by the weaker increase in activation for uncertain gains. It is also consistent with the difference in PA for losses (but not gains) across incentive sets, since NAcc activation correlates with PA (Knutson and Greer, 2008). However, other factors might

also explain this pattern. For example, expected value for uncertain losses is closer to zero than for uncertain gains (about  $-\$1.67$  vs  $\$3.33$ ), and the influence of available alternative incentives might be larger for small incentives. As well, activation for uncertain gains is already relatively high in the UNC-ONLY set, and so physiological ceiling effects might also more simply account for the weaker change for gains. The current design cannot distinguish between these accounts, but further research should explore how incentive magnitude in particular might affect the influence of other available incentives.

The findings have a final implication for interpreting the neural representations of value during decisions with multiple available incentives, especially mixed certain and uncertain incentives (as is the case for many gambles). The current results suggest that the availability of alternative certain incentives can influence the neural representations of uncertain incentives with similar magnitude. Representations of an uncertain incentives in a decision involving certain incentives, then, might differ from representations of that same incentive in a decision involving only uncertain incentives. This argues for a note of caution when comparing decisions with different kinds of incentives; the mere availability of certain incentives may qualitatively change how the brain represents the value of uncertain incentives.

The present design controlled for several potential confounds between contexts, but also had some limitations. One important limitation is the fixed inter-trial interval, which raises the possibility that the end of the previous trial might have influenced the time course of activation in response to a cue (see Supplementary Figure S2 for full time course data). This may have decreased NAcc activation at the beginning of each trial, so that most of our measures of activation began below the experimental mean. The initial decrease makes it difficult to identify the onset of contextual differences in time. However, the decrease did not vary by condition or incentive set, perhaps since anticipatory NAcc activation is not strongly influenced by prior outcomes in the MID task (Knutson *et al.*, 2001b). The decrease in NAcc activation at trial onset is consistent with earlier studies (Larkin *et al.*, 2007), suggesting that the main conclusions were not distorted by potential activation carryover from trial to trial.

In summary, this study demonstrated that in the NAcc, anticipatory representations of an incentive can be modulated by the mere availability of alternative incentives. Specifically, anticipatory NAcc activation for uncertain losses (and to a lesser extent gains) increased when both certain gains and losses were available. The pattern of effects was most consistent with the idea that participants used certain losses (but not gains) as an anchor when they were available. This anchor account highlights the flexible nature of value representations. Neural systems can alter how they represent the environment from second to second. Understanding and predicting these changes requires

innovative experimental designs and incremental research. One important future direction might involve varying the available incentives in an event-related fashion, to capture how representations might change from trial to trial. To further understand how the brain anticipates incentives in a changing world, studies may profitably use designs and analyses that honor its dynamic flexibility.

## SUPPLEMENTARY DATA

Supplementary data are available at SCAN online.

## Conflict of Interest

None declared.

## REFERENCES

- Bayer, H.M., Glimcher, P.W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron*, 47, 129–41.
- Bjork, J.M., Knutson, B., Fong, G.W., Caggiano, D.M., Bennett, S.M., Hommer, D.W. (2004). Incentive-elicited brain activation in adolescents: similarities and differences from young adults. *Journal of Neuroscience*, 24, 1793–802.
- Bjork, J. M., Hommer, D.W. (2007). Anticipating instrumentally obtained and passively-received rewards: a factorial fMRI investigation. *Behavioural Brain Research*, 177, 165–70.
- 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, 619–39.
- Breiter, H.C., Gollub, R.L., Weisskoff, R.M., et al. (1997). Acute effects of cocaine on human brain activity and emotion. *Neuron*, 19, 591–611.
- Brett, M., Anton, J.-L., Valabregue, R., Poline, J.-B. (2002). Region of interest analysis using an SPM toolbox (abstract). Presented at 8th International Conference on Functional Mapping of the Human Brain. Sendai, Japan.
- Cooper, J.C., Knutson, B. (2008). Valence and salience contribute to nucleus accumbens activation. *NeuroImage*, 39, 538–47.
- Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance images. *Computers in Biomedical Research*, 29, 162–73.
- Devlin, J.T., Poldrack, R. A. (2007). In praise of tedious anatomy. *NeuroImage*, 37, 1033–41.
- Elliott, R., Agnew, Z., Deakin, J.F. (2008). Medial orbitofrontal cortex codes relative rather than absolute value of financial rewards in humans. *European Journal of Neuroscience*, 27, 2213–18.
- Glover, G.H., Law, C.S. (2001). Spiral-in/out BOLD fMRI for increased SNR and reduced susceptibility artifacts. *Magnetic Resonance in Medicine*, 46, 515–22.
- Guyer, A.E., Nelson, E.E., Perez-Edgar, K., et al. (2006). Striatal functional alteration in adolescents characterized by early childhood behavioral inhibition. *Journal of Neuroscience*, 26, 6399–405.
- Ikemoto, S., Panksepp, J. (1999). The role of nucleus accumbens dopamine in motivated behavior: a unifying interpretation with special reference to reward-seeking. *Brain Research Reviews*, 31, 6–41.
- Knutson, B., Cooper, J.C. (2005). Functional magnetic resonance imaging of reward prediction. *Current Opinion in Neurology*, 18, 411–17.
- Knutson, B., Greer, S.M. (2008). Anticipatory affect: neural correlates and consequences for choice. *Philosophical Transactions of the Royal Society B*, 363, 3771–86.
- Knutson, B., Peterson, R. (2005). Neurally reconstructing expected utility. *Games and Economic Behavior*, 52, 305–15.
- Knutson, B., Adams, C.M., Fong, G.W., Hommer, D. (2001a). Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *Journal of Neuroscience*, 21, RC159.
- Knutson, B., Fong, G.W., Adams, C.M., Varner, J.L., Hommer, D. (2001b). Dissociation of reward anticipation and outcome with event-related fMRI. *NeuroReport*, 12, 3683–7.
- Knutson, B., Fong, G.W., Bennett, S.M., Adams, C.M., Hommer, D. (2003). A region of mesial prefrontal cortex tracks monetarily rewarding outcomes: characterization with rapid event-related fMRI. *NeuroImage*, 18, 263–72.
- Knutson, B., Rick, S., Wimmer, G.E., Prelec, D., Loewenstein, G. (2007). Neural predictors of purchases. *Neuron*, 53, 147–56.
- Knutson, B., Taylor, J., Kaufman, M.T., Peterson, R., Glover, G. (2005). Distributed neural representation of expected value. *Journal of Neuroscience*, 25, 4806–12.
- Kriegeskorte, N., Simmons, W.K., Bellgowan, P.S.F., Baker, C. I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. *Nature Neuroscience*, 12, 535–40.
- Larkin, G.R., Gibbs, S.E., Khanna, K., Nielsen, L., Carstensen, L.L., Knutson, B. (2007). Anticipation of monetary gain but not loss in healthy older adults. *Nature Neuroscience*, 10, 787–91.
- Luce, R.D. (1992). Where does subjective expected utility fail descriptively? *Journal of Risk and Uncertainty*, 5, 5–27.
- Montague, P.R., King-Casas, B., Cohen, J.D. (2006). Imaging valuation models in human choice. *Annual Review of Neuroscience*, 29, 417–48.
- Nieuwenhuis, S., Heslenfeld, D.J., von Geusau, N.J., Mars, R.B., Holroyd, C.B., Yeung, N. (2005). Activity in human reward-sensitive brain areas is strongly context dependent. *NeuroImage*, 25, 1302–9.
- O'Doherty, J.P. (2004). Reward representations and reward-related learning in the human brain: insights from neuroimaging. *Current Opinion in Neurobiology*, 14, 769–76.
- Preston, A.R., Thomason, M.E., Ochsner, K.N., Cooper, J.C., Glover, G.H. (2004). Comparison of spiral-in/out and spiral-out BOLD fMRI at 1.5 and 3 T. *NeuroImage*, 21, 291–301.
- Preusschoff, K., Bossaerts, P., Quartz, S. R. (2006). Neural differentiation of expected reward and risk in human subcortical structures. *Neuron*, 51, 381–90.
- Russell, J.A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39, 1161–78.
- Schultz, W. (2006). Behavioral theories and the neurophysiology of reward. *Annual Review of Psychology*, 57, 87–115.
- 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, 4826–31.
- Simonson, I., Tversky, A. (1992). Choice in context: tradeoff contrast and extremeness aversion. *Journal of Marketing Research*, 29, 231–95.
- Tobler, P.N., Fiorillo, C.D., Schultz, W. (2005). Adaptive coding of reward value by dopamine neurons. *Science*, 307, 1642–5.
- 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, 515–18.
- Tversky, A., Kahneman, D. (1992). Advances in prospect theory: cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5, 297–323.
- Tversky, A., Simonson, I. (1993). Context-dependent preferences. *Management Science*, 39, 1179–89.
- Ward, B.D. (2002). AlphaSim. <http://afni.nimh.nih.gov>.
- Watson, D., Wiese, D., Vaidya, J., Tellegen, A. (1999). The two general activation systems of affect: structural findings, evolutionary considerations, and psychobiological evidence. *Journal of Personality and Social Psychology*, 76, 820–38.