

Neural antecedents of social decision-making in a partner choice task

Samuel C. D. Cartmell,^{1,2} Marvin M. Chun,¹ and Timothy J. Vickery^{1,3}

¹Department of Psychology, Yale University, PO Box 208205, 2 Hillhouse Ave, New Haven, CT 06520-8205, ²Stanford School of Medicine, 291 Campus Dr, Stanford, CA 94305, and ³Department of Psychology, University of Delaware, 108 Wolf Hall, Newark, DE 19716

Experiments in financial decision-making point to two complementary processes that encode prospective gain and loss preceding the choice to purchase consumer goods. These processes involve the nucleus accumbens (NAcc) and the right anterior insula, respectively. The current experiment used functional MRI to investigate whether these regions served a similar function during an analogous social decision-making task without the influence of monetary outcomes. In this task, subjects chose partners based on face stimuli of varying attractiveness (operationalizing value) and ratings of compatibility with the participant (operationalizing likelihood of rejection). The NAcc responded to anticipated gain; the right anterior insula responded to compatibility, but not in a manner that suggests an analogy to anticipated cost. Logistic regression modeling demonstrated that both regions predicted subsequent choice above and beyond the influence of group attractiveness ratings or compatibility alone. Although the function of the insula may differ between tasks, these results suggest that financial and social decision-making recruit a similar network of brain regions.

Keywords: social decision making; reward; attractiveness; nucleus accumbens

INTRODUCTION

Recent evidence suggests that two distinct and complementary neural systems underlie financial decision-making. Although one network, centered in the ventral striatum, computes anticipated gain (Knutson *et al.*, 2001; Kable and Glimcher, 2007; Lebreton *et al.*, 2009; Levy *et al.*, 2011), the other, involving the anterior insula, tracks anticipated cost, as evidenced by the insula's response to price during a goods purchasing task (Knutson *et al.*, 2007) and its role in the anticipation of aversive financial outcomes (Paulus *et al.*, 2003; Kuhnen and Knutson, 2005; Paulus and Stein, 2006; Samanez-Larkin *et al.*, 2008). Furthermore, the anterior insula is likely involved in representing the closely related concept of anticipated risk (Knutson and Bossaerts, 2007; Mohr *et al.*, 2010).

However, it remains unresolved whether these same neural systems predict choice in other domains, such as social decision-making. Recent work is beginning to link financial decision-making and social decision-making; for example, the ventral striatum is activated by social rewards such as praise, and the anterior insula by aversive social outcomes such as a partner's defection during a prisoner's dilemma (PD) game (Izuma *et al.*, 2008; Lee, 2008; Rilling *et al.*, 2008, Rilling and Sanfey, 2011). A common system may underlie both types of decisions, but evidence linking financial and social decision-making remains limited, as choices in social decision-making tasks typically still confound monetary incentives with social incentives (e.g. the PD game, where the goal is to maximize the amount of money won).

The current functional MRI (fMRI) experiment explores whether anticipated social gain and social cost evoke analogous activity in brain regions shown to code financial gain and cost in the aforementioned studies. Our task structure is inspired by a financial decision task used to reveal anticipatory networks related to the purchasing of

different goods—specifically, the Save Holdings or Purchase (SHOP) paradigm (Knutson *et al.*, 2007). In the original SHOP paradigm, subjects are presented with a picture of an item for a short period, then the price is displayed alongside the item, and finally subjects are prompted to decide if they would like to purchase the item at the given price. We modified this paradigm so that instead of making decisions about the purchase of goods based on preference and cost, subjects were asked to choose potential romantic partners based on the target's appearance (attractiveness), corresponding to preference, and incompatibility in personality (between the target and subject). Incompatibility was directly linked to the likelihood of rejection in this experiment, and thus represented a social cost (see also General Discussion). Although social decisions are typically studied in the context of games which may compare conditions with equivalent financial outcomes but nonetheless still involve money, our task did not require any decisions that involved financial reward in any way (beyond the same base compensation that everyone received regardless of performance). Hence, incentives in this task are purely 'social'.

If attractiveness in social decision-making is analogous to preference in financial decision-making, then activity centered in mesolimbic dopamine projections in the ventral striatum, especially the nucleus accumbens (NAcc), should increase with increasing attractiveness. And if social incompatibility is treated as a cost, then anterior insula activity should increase with increasing incompatibility between the target and the subject. As our task was designed to be a social analogue of Knutson *et al.*'s (2007) financial SHOP paradigm, our analysis of brain activity focused primarily on the a priori regions of the NAcc and right anterior insula. We excluded consideration of the MPFC, thought to be important for tracking the price differential in Knutson *et al.* (2007), as an a priori region of interest (ROI) because our study did not contain an analogous metric (i.e. a difference between how compatible subjects thought the target was and the rating that was actually presented). We further use logistic regression analyses to explore the link between activity in these regions and subsequent choice. Finally, we examine our results with *post hoc* whole brain analyses. Behaviorally, choice of a target should be increased by attractiveness and decreased by incompatibility, as these dimensions respectively index the choice's potential benefit and likelihood of rejection.

Received 27 July 2012; Revised 22 October 2013; Accepted 31 October 2013

Advance Access publication 5 November 2013

This work was supported by the Kavli Institute for Neuroscience at Yale, a gift from the Yudofsky family, and the FAS MRI program by the Office of the Provost and Department of Psychology at Yale University.

Correspondence should be addressed to Samuel C. D. Cartmell, 24 Roble Ct, Berkeley, CA 94705, USA. E-mail: samcartmell@gmail.com

MATERIALS AND METHODS

Participants

The fMRI study recruited 19 self-identified heterosexual volunteers from the Yale community (12 female; right-handed; mean age 20) who were not in a monogamous relationship at the time of the study. Three subjects were excluded from the analysis—one due to excessive motion (>7 mm movement within one run), one for failing to complete the final block and one for technical errors during the scanning session that rendered the data unusable. Separately, a survey was conducted from an additional 127 participants who rated the physical attractiveness of the faces used in this experiment.

Materials

A total of 192 images, 96 from each gender, were used in the final version of the experiment. These images were obtained from Internet searches, the NimStim database (Tottenham *et al.*, 2009) and the PAL aging database (Minear and Park, 2004) ages 18–29 subgroup. To separate high and low attractiveness images, a set of 233 images (115 male, 118 female) were rated for physical attractiveness on a scale from 1 to 7. Male and female images were viewed by a separate set of undergraduate participants who rated the images of the gender that they identified as most attractive. The 48 highest and lowest rated images from each gender were used as the high and low attractiveness images, respectively, whereas images whose rating fell in between these groups were discarded (male images mean high rating = 3.49, *s.d.* = 0.84, mean low = 2.10, *s.d.* = 0.21; female images mean high 4.77, *s.d.* = 0.84, mean low = 2.31, *s.d.* = 0.46). The personality test was based on an assessment used by Chemistry.com to evaluate personality, and was modified for length from 82 questions to 35. The task was administered through a computer program coded in the MATLAB language, using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997).

Imaging procedure

Structural and functional data were acquired on a 3T Siemens Trio system with a 12-channel head coil. Sagittal T1-weighted MPRAGE images with 176 slices and 1 mm isotropic voxels were used for cortical and subcortical labeling and participant coregistration. Functional scans were T2*-weighted gradient-echo EPI sequences, consisting of 34 slices with an oblique axial orientation and 3.5 mm × 3.5 mm in-plane resolution, 4 mm thickness, 0 mm spacing between slices (sequence parameters: TR = 2000 ms, TE = 25 ms, FA = 90°, matrix: 64 × 64). Four functional scanning runs consisting of 279 volumes including five discarded volumes were acquired for each participant, with each run lasting 9 min 18 s.

Task design

There were eight experimental conditions, reflecting a 2 × 2 × 2 (order × attractiveness × compatibility) design. Each trial presented face and compatibility information, and then required subjects to indicate whether they were romantically interested in the person presented. The face image could be either attractive or unattractive, and the compatibility rating could either be low or high as described below. In order to separately estimate the hypothesized effects of compatibility and attractiveness, in half the trials face information was presented first and compatibility second, and the sequence was reversed in the other half. In the face first trials, the face was presented alone for 6 s before the rating was presented alongside it for another 6 s. In the compatibility first trials, the timing was identical but the sequence was reversed. The primary task consisted of four runs of 24 trials, for a total of 96 trials per subject.

Procedure

Subjects were recruited to participate in a study titled ‘Attraction and Partner Selection’ and given a brief description that they would be required to supply personality information and make decisions about the desirability of other people. Subjects were also told that the personality information they provided would be seen by another group of participants, and that this second group would rate their level of interest based on the answers the subjects gave. Upon sign-up, subjects were asked to immediately complete an online personality survey. After at least 1 week following completion of the survey, subjects came into the laboratory to complete the task.

On each trial of the primary task, subjects were presented with an image of a smiling face of the opposite gender and a ‘compatibility’ rating, and were then required to indicate whether or not they were romantically interested in this person on the basis of the image and rating. In an attempt to create a more ecologically valid situation, and with IRB approval, subjects were told that the images were headshots of other participants in the study who had also filled out the personality survey, and that these other participants had seen the subject’s responses to the personality survey. In addition, subjects were told that the other participants had indicated whether they were romantically interested in the subject based on the subject’s responses to this survey. Finally, subjects were told the compatibility ratings were a number from a scale of 0–10, generated by an algorithm that computed the degree of overall fit or compatibility between the personality of the subject and that of the other participant based on each party’s answers to the personality survey. As they would in a real social situation, subjects were instructed to ‘consider the other party’s reaction’ when making their choice by attending to the compatibility rating, which they were told was a ‘good indication’ of how the other participant was likely to view the subject. Subjects only saw faces of the opposite gender and were told no information about the potential to meet the other participant(s) after the experiment. In reality, the depicted individuals were not part of the experiment, and images were pulled from the Internet. Compatibility ratings could be low or high depending on the experimental condition—high compatibility ratings were randomly selected from the set {7, 8, 9}, whereas low ratings were randomly selected from the set {1, 2, 3}. The target’s responses were dependent on the compatibility rating in a probabilistic manner (compatibility rating ÷ 10 = the probability of a ‘Yes’ response), so that a compatibility rating of 8, for instance, would correspond to an 80% chance of the target responding ‘Yes’.

Trials were broken into five stages (Figure 1a). In half the trials, subjects were first presented with a face for 6 s, after which the compatibility rating appeared alongside the face for 6 s (face first condition). For the other randomly intermixed trials, the presentation order of the rating and face was switched so that the rating appeared prior to the face (compatibility first condition). In the third stage of all trials, the face and rating remained onscreen, while subjects made their choice by selecting either ‘Yes’ or ‘No’, which appeared on either side of the face image. In the fourth stage, feedback indicating the other participant’s choice was presented for 2 s, which, depending on the compatibility rating, displayed either ‘Interested’ (if the subject chose ‘No’) ‘Not Interested in You’ (regardless of the subject’s choice) or ‘Match!’ (if both the subject’s and target’s response was ‘Yes’). In the fifth and final stage, fixation was displayed for 4 s before the next trial commenced. In all trials, subjects used either the index or middle finger of their right hand to select a response on a button box. The index finger always chose the option on the left; the middle finger always chose the option on the right, and the options (‘Yes’ and ‘No’) switched position randomly on each trial.

After the task was complete, subjects were debriefed and all aspects of the experiment and deception were fully explained to them. Upon

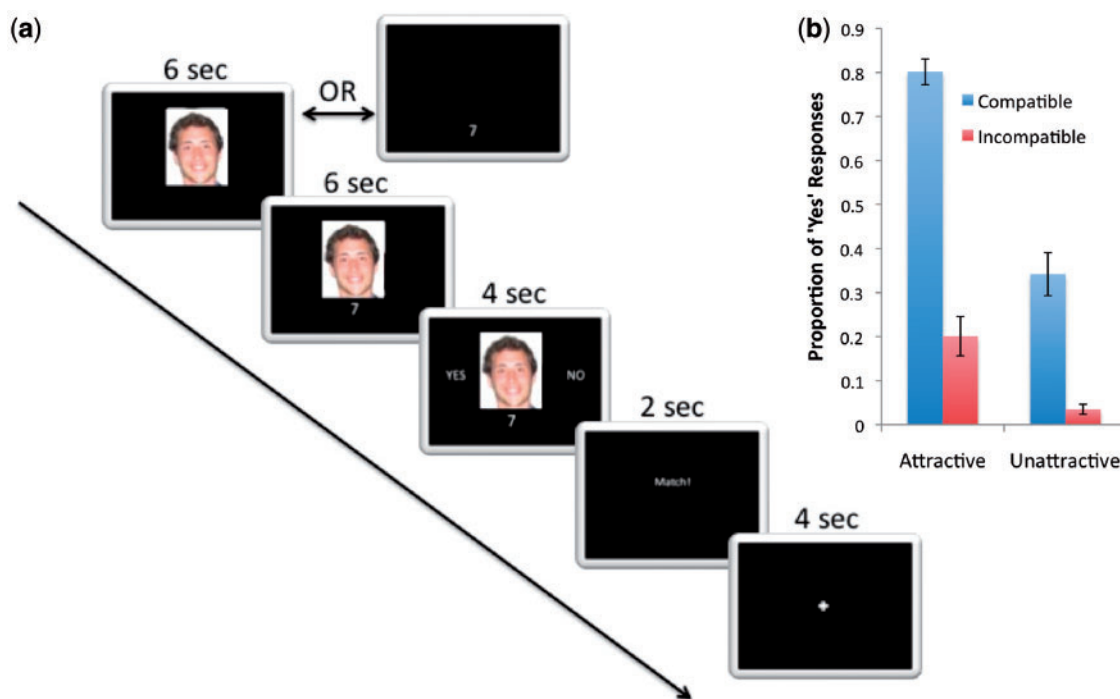


Fig. 1 Task trial structure and behavioral results showing the proportion of 'Yes' responses associated with each condition.

debriefing, no subjects reported they felt suspicious of the limited range of compatibility ratings, or that they had serious doubts about the cover story.

fMRI data analysis

Analysis and preprocessing of fMRI data were conducted with the FreeSurfer software package (<http://surfer.nmr.mgh.harvard.edu>). Functional data were slice-time corrected and motion-corrected to the first volume of the first functional scan. The functional data were then registered with each subject's high-resolution anatomical dataset, spatially smoothed with a 4 mm FWHM Gaussian kernel, and normalized to a standard template in MNI305 space. Beta coefficients were estimated using a General Linear Model with a polynomial baseline fit, and motion parameters and global signal as covariates of no interest.

To test attractiveness, compatibility, and order effects, the fMRI data were analyzed in a Finite Impulse Response (FIR) model where each trial was assigned one of eight conditions ($2 \times 2 \times 2$, attractiveness \times compatibility \times order), and each of the nine TRs within the 18 s trial was modeled separately, generating a beta weight for each condition for each timepoint. In the Yes > No analysis, a similar approach was used to create another model, except that each trial was instead assigned one of four conditions (2×2 , order \times choice). This four-condition analysis was chosen in place of a 16-condition matrix ($2 \times 2 \times 2 \times 2$, attractiveness \times compatibility \times order \times choice) due to the fact that this sometimes resulted in empty bins—certain conditions in some subjects had no 'Yes' choices. Contrasts were computed between the beta weights of different conditions at specific timepoints. Of particular interest was the activity 4–8 s after each stimulus presentation, as this is when stimulus-related activity would peak according to a canonical hemodynamic response function. As there were two stimulus onsets in each trial, analysis focused on the averaged activity 4–8 s after each onset, hereafter labeled Interval 1 and Interval 2. During Interval 1 either the face or compatibility rating was visible (but not both), whereas during Interval 2, both the image and rating were on

the screen. Activity during these periods precedes activity related to the choice phase in which subjects made their decision.

Based on Knutson *et al.* (2007), initial fMRI data analysis focused on a priori regions of the right anterior insula and bilateral NAcc, defined anatomically using the Destrieux atlas from FreeSurfer's automated parcellations (Fischl *et al.*, 2004; Destrieux *et al.*, 2010). Because FreeSurfer's original anterior insula sub-parcellation produced an ROI with limited extent (mean < 20 functional voxels), this anterior insula region was expanded to include all adjacent 1 mm³ anatomical voxels, ultimately resulting in larger ROIs (mean = 78.7 functional voxels). The mean activity during each timepoint for each condition was then extracted from these anatomically defined regions.

To test whether inclusion of activity from the a priori regions improved prediction, we implemented competing logistic regression models. Both models used targets' attractiveness and compatibility information (coded as binary variables) to predict choice on a trial-by-trial basis; however, one model also incorporated data from the bilateral NAcc and right anterior insula during Interval 2 as separate regressors. Each model's maximum likelihood estimate (MLE) was calculated as the product of each subject's MLE for that model, thus pooling all subjects and trials. χ^2 statistics obtained from the likelihood ratio test served as the basis for model comparison. Three subjects were excluded from this analysis because their data were perfectly separated using only the independent variables (so no valid MLE could be estimated), although inclusion of these subjects did not affect the pattern of results or conclusions. For completeness, implementation of these models was repeated using continuously varying measures (as opposed to the binary coding used above) of attractiveness and compatibility (which were only available for 12/16 subjects). The pattern of results and subsequent conclusions in this model were identical; thus, only results from the first model are presented for concision.

We further conducted exploratory *post hoc* whole brain analyses to identify regions that predicted choice or responded to either of the independent variables of compatibility and attractiveness. This secondary analysis considered activations which survived a voxelwise

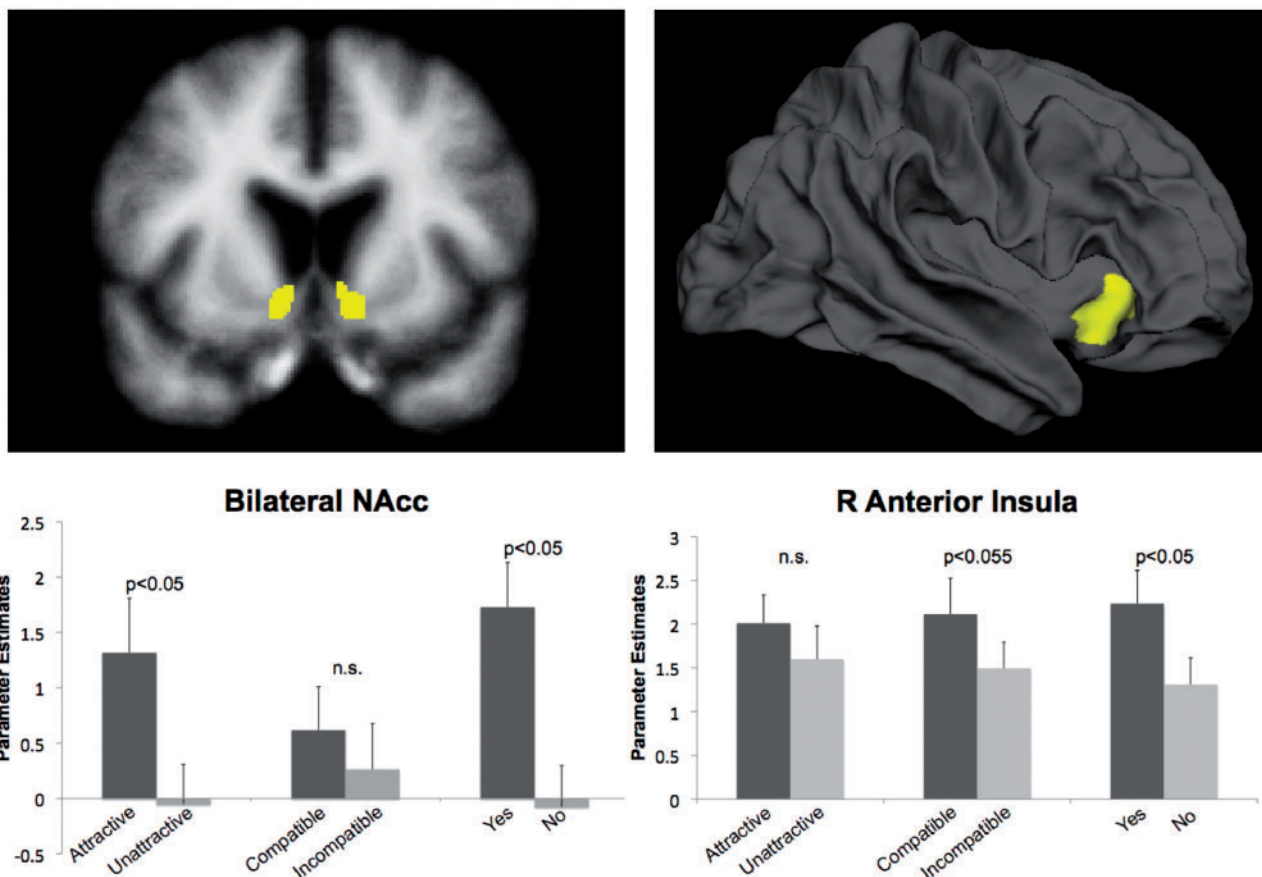


Fig. 2 (Bottom) Activity of the a priori ROIs from Interval 2. (Top) Brain images showing the anatomically defined ROIs created with FreeSurfer’s parcellation algorithm. The Yes/No and Compatibility/Attractiveness parameter estimates are drawn from separate FIR models (see Methods: fMRI Data Analysis for details). Error bars represent standard error of the mean.

threshold of $P < 0.005$ and a clusterwise probability of $P < 0.05$ based on Gaussian random field correction for multiple comparisons (Worsley *et al.*, 1992; Cho *et al.*, 2010).

RESULTS

Behavioral results

Figure 1b shows the behavioral data. A 2 (order) \times 2 (attractiveness) \times 2 (compatibility) repeated measures analysis of variance (ANOVA) was conducted on the remaining 16 subjects’ behavioral responses during scanning. The probability of a subject choosing ‘Yes’ on a given trial was used as the dependent measure. As order was of secondary interest, and the main effect and interactions involving order were not statistically significant, only effects related to attractiveness and compatibility are reported.

The ANOVA revealed a main effect of attractiveness on choice, $F(1, 15) = 121.12, P < 0.0005$, and a main effect of compatibility $F(1, 15) = 55.98, P < 0.0005$, with both more attractive and more compatible candidates yielding more ‘Yes’ responses. There was a significant interaction of attractiveness and compatibility, $F(1, 15) = 28.49, P < 0.0005$, such that the effect of compatibility is attenuated in low attractiveness targets. There was no significant difference in reaction time between ‘Yes’ and ‘No’ responses [paired two-tailed t -test $t(15) = 1.15, P = 0.27$].

fMRI results

ROI analysis

Interval 1. To examine attractiveness-related activity in a priori regions, a one-way ANOVA was performed on extracted activity

from Interval 1 for face first trials only. Consistent with hypotheses, the bilateral NAcc showed significantly greater activity for attractive rather than unattractive trials $F(1, 15) = 9.55, P < 0.01$. The right anterior insula did not show a significant response to attractiveness $F(1, 15) = 2.13, P = 0.16$.

Compatibility effects were similarly examined by performing a one-way ANOVA on Interval 1 activity for compatibility first trials only. Neither a priori region showed a significant response to compatibility during this period.

Finally, for completeness, the ability of these regions to predict choice during Interval 1 was examined by performing a 2 \times 2 ANOVA (order \times choice) on activity collapsed across both compatibility and face first trials. Neither region showed a significant main effect of choice nor was there a significant interaction in either region; however, the right anterior insula showed a main effect of order such that if a face was displayed during Interval 1, its activity increased $F(1, 15) = 31.801, P < 0.001$.

Interval 2. A 2 \times 2 \times 2 (order \times attractiveness \times compatibility) ANOVA was also performed for each region on the Interval 2 data, during which both attractiveness and compatibility information are available. Figure 2 summarizes the Interval 2 data for these regions. Bilateral NAcc showed greater activity for attractive rather than unattractive trials, $F(1, 15) = 8.43, P < 0.05$, and no significant main effects or interactions of order or compatibility.

The right anterior insula showed a marginally significant main effect of compatibility $F(1, 15) = 4.35, P = 0.05$, but, in contrast to our hypothesis, activity was greater for compatible rather than incompatible trials. The insula showed no other significant main effects or interactions.

Table 1 Attractiveness regions

Region		Talairach coordinates			z-score	Clusterwise <i>P</i> -value
		X	Y	Z		
Interval 1	Attractive > unattractive face first					
R/L mOFC		−10	29	−14	4.37	<0.0001
R/L NAcc		2	4	−11	4.3	<0.0001
L lingual		−20	−60	−6	3.94	0.0332
R lingual		16	−75	−2	3.91	<0.0001
L cuneus		−2	−91	28	3.79	0.0227
Interval 2	Attractive > unattractive (collapsed)					
R/L superior frontal		−4	8	54	4	0.0309
L precentral		−42	3	38	3.79	0.0327
L anterior insula		−34	7	1	3.67	0.0259
Interval 2	Attractive > unattractive face first					
n/a						
Interval 2	Attractive > unattractive compatibility first					
L caudal MFG		−27	13	41	4.36	0.0425

Both the NAcc and the right anterior insula predicted choice, showing significantly greater activity during 'Yes' vs 'No' trials in this period [insula $F(1, 15) = 9.92$, $P < 0.01$, NAcc $F(1, 15) = 10.962$, $P < 0.01$]. There were no significant interactions of order and choice in either region during this interval.

Logistic regression

Two logistic regression models were implemented to establish whether the addition of brain activity to the model improved prediction over the use of the independent variables alone. The first model regressed subjects' choices on compatibility and attractiveness information, whereas the second model regressed subjects' choices on the independent variables and neural data from a priori regions immediately following presentation of the appropriate stimulus; the models were compared using the likelihood ratio test. The results of this comparison showed that the addition of the brain activity substantially improved the fit of the model (Table 3). Specifically, the likelihood ratio test yielded a highly significant χ^2 statistic (26, $N = 26$) = 54.3 ($P < 10^{-3}$), favoring the brain model. Other measures of model fit such as adjusted (McFadden) pseudo- R^2 and Akaike information criterion (AIC) similarly select for the brain model. These results are unlikely to be explained by overfitting as both AIC and pseudo- R^2 values include a penalty for additional parameters.

All parameters except insula activity were reliably above zero in both models. This is likely due to the fact that neural data were taken from both Interval 1 and 2 to make predictions, and significant compatibility-related insula activity was not observed until the second interval regardless of stimulus presentation order. This suggests that ratings are not meaningful in the absence of a face. To further explore the utility of the insula in predicting choice, a separate logistic regression model using data only from Interval 2 (for both regions) was implemented (Table 3, model 3). The insula becomes a significant predictor, although the NAcc is no longer statistically significant. This model substantially improves fit compared with both the IV-only model and the previously described brain model in all measures of model fit. Both models 2 and 3 demonstrate that incorporation of brain activity improves prediction above and beyond the use of the independent variables alone.

Whole brain

Each of the following contrasts was bidirectional; however, many contrasts only had significant activity in one direction—e.g.

compatible > incompatible. Therefore, if no regions are reported for a given direction it is due to the fact that none were significant. In addition, the reader should note that the following Yes vs No contrasts are not statistically independent from the effects of compatibility and attractiveness. Nonetheless, the large overlap between regions that differentiate between subsequent choice and those that respond to compatibility and attractiveness, as well as the results of the logistic regression, suggests that these regions represent anticipated loss, anticipated gain and subsequent choice.

Interval 1. To explore the regions that preferentially responded to attractiveness, a whole brain contrast of High vs Low attractiveness trials ([AttrComp + AttrIncomp] − [UnattrComp + UnattrIncomp]) was conducted for the face first Interval 1 data (Table 1). Robust activations of the bilateral NAcc and bilateral medial orbitofrontal cortex (mOFC) were observed, consistent with other studies that compared attractive vs unattractive faces (O'Doherty et al., 2003; Winston et al., 2007).

To explore the regions that preferentially responded to compatibility, a whole brain contrast of High vs Low compatibility trials ([AttrComp + UnattrComp] − [AttrIncomp + UnattrIncomp]) was conducted on compatibility first Interval 1 data. There were no regions that survived correction for multiple comparisons during compatibility first Interval 1.

To explore regions that predicted choice, a whole brain contrast of 'Yes' vs 'No' trials was conducted separately on the compatibility first and face first Interval 1 data. The Interval 1 face first contrast showed several significant clusters, most notably in the bilateral mOFC, bilateral caudal Anterior Cingulate Cortex (ACC) and bilateral caudal middle frontal gyrus. The activations here likely reflect attractiveness-related activity, particularly in the mOFC, which was present in the attractive vs unattractive contrast. The Interval 1 compatibility first contrast showed no significant regions.

Interval 2. During Interval 2, the bilateral superior frontal gyrus, left precentral gyrus and left anterior insula responded significantly to attractiveness. Bilateral rostral ACC, bilateral superior frontal gyrus and bilateral caudate all showed greater activation for compatible rather than incompatible conditions. A whole brain contrast of the interaction between attractiveness and compatibility ([AttrComp + UnattrIncomp] − [AttrIncomp + UnattrComp]) revealed bilateral activations in the mOFC, indicating that these frontal regions are

Table 2 Compatibility regions and interaction

Region		Talairach coordinates			z-score	Clusterwise P-value
		X	Y	Z		
Interval 1	Compatible > incompatible compatibility first					
n/a						
Interval 2	Compatible > incompatible (collapsed)					
R/L rostral ACC, superior frontal		-6	35	14	5.02	<0.0001
L cerebellum		-44	-69	-33	4.91	0.004
L caudate		-8	18	-2	4.59	<0.0001
R caudate		12	11	-3	4.22	0.0001
R lateral occipital		34	-81	-4	4.11	<0.0001
R cerebellum		32	-61	-36	3.94	0.0248
L inferior temporal		-48	-66	-9	3.78	0.0352
L lateral occipital		-22	-88	4	3.66	<0.0001
Interval 2	Compatible > incompatible face first					
R cerebellum		32	-63	-35	4.77	0.0003
R/L superior frontal		6	29	25	4.62	<0.0001
L superior frontal		-8	35	14	4.33	0.0422
R lateral occipital		40	-66	-11	4.2	0.0002
R superior parietal		28	-65	34	4.03	0.0013
R/L PCC		0	-34	36	3.74	0.0029
L lateral occipital		-42	-87	-3	3.7	0.0003
Interval 2	Compatible > incompatible compatibility first					
L rostral ACC		-8	39	16	4.56	0.0013
R/L caudate		-10	11	1	4.35	<0.0001
R superior frontal		4	15	39	3.86	0.0005
Interval 2	Attractiveness × compatibility (collapsed)					
L mOFC		-8	27	-17	4.65	0.0006
R mOFC		8	27	-17	4.01	0.0078

Table 3 Comparison of logistic regression models

	(1) Independent variables	(2) Combined	(3) Combined, Int 2 only
Constant	-3.74 ^a s.d. = 1.46	-3.87 ^a s.d. = 1.42	-4.06 ^a s.d. = 1.59
Attractiveness	2.37 ^a s.d. = 1.1	2.4 ^a s.d. = 1.13	2.51 ^a s.d. = 1.24
Compatibility	2.77 ^a s.d. = 1.59	2.83 ^a s.d. = 1.64	2.87 ^a s.d. = 1.72
R/L NAcc		0.83 ^a s.d. = 1.15	0.42 s.d. = 1.55
R insula		0.42 s.d. = 1.13	1.27 ^a s.d. = 1.62
χ ² statistic ^b	n/a	(26, N = 26) = 54.3 P < 10 ⁻³	(26, N = 26) = 69.67 P < 10 ⁻⁵
AIC	1086.2	1083.9	1068.5
Adj. pseudo-R ²	0.4234	0.4247	0.4328

^aSignificantly different from zero, P < 0.05, two-tailed t-test. ^bBased on likelihood ratio test relative to model (1).

sensitive to both factors. To further characterize the response in these regions, we created a mask containing all the significant functional voxels from each activation in the interaction contrast (L and R mOFC) and extracted the activity. Both L and R mOFC show the same pattern, such that the BOLD response was modulated by attractiveness in the compatible condition and not the incompatible condition (mean beta weights, collapsed bilaterally: attractive-compatible = 1.02, unattractive-compatible = -0.01, attractive-incompatible = 0.52, unattractive-incompatible = 0.54). **Table 2** reports more detailed analyses.

Areas that predicted subsequent choice during Interval 2 spanned a large and robust network of regions whose activity was greater in ‘Yes’ than ‘No’ trials. This circuit included numerous regions that have been

previously implicated in decision-making tasks, including the bilateral ventral striatum, bilateral mOFC, left lateral OFC, bilateral ACC, bilateral Posterior Cingulate Cortex (PCC), bilateral thalamus and bilateral superior frontal gyrus (Manes *et al.*, 2002; Knutson *et al.*, 2007; Lee, 2008; Rilling and Sanfey, 2011). The anterior insula strongly predicted choice bilaterally, although contrary to expectations greater activation was observed for ‘Yes’ rather than ‘No’ trials (see **Supplementary Data**).

When broken down by order, the compatibility first Interval 2 data reconstituted nearly the whole frontal and subcortical network found in the Yes > No contrast collapsed across order. In particular, activation of the bilateral ventral striatum, bilateral mOFC, bilateral thalamus, bilateral superior frontal gyrus, bilateral rostral ACC, bilateral PCC and left lateral OFC was observed. The face first contrast included many of the same regions, including bilateral rostral ACC, bilateral superior frontal, bilateral PCC, left lateral OFC, left caudate and thalamus.

Finally, to determine which regions showed effects due to the order of stimulus presentation, a whole brain contrast comparing face first conditions to compatibility first conditions was conducted on the Interval 2 data. The full results are listed in the **Supplementary Data**—these activations likely reflect processing of the most recently displayed stimulus (e.g. bilateral fusiform activation for compatibility first > face first trials) rather than any behaviorally relevant activity.

DISCUSSION

This study investigated the neural substrates of social decision-making, especially as they relate to regions previously shown to mediate choice and decision-making in financial tasks. Accordingly, to facilitate comparison, the task mirrored previous financial decision-making paradigms while substituting social analogues of preference and cost—in this case targets’ attractiveness and incompatibility, respectively.

Attractiveness is known to be a sought-after reward, and compatibility ratings here signaled whether the target person would reciprocate interest in the subject. Behaviorally, we hypothesized main effects of compatibility and attractiveness and a significant interaction. Neurally, we predicted from the financial decision-making literature that the NAcc should encode prospective benefit, the right anterior insula should reveal prospective loss, and both of these areas should predict choice (Knutson *et al.*, 2007).

As hypothesized, both attractiveness and compatibility showed a main effect on choice behavior, and an interaction, such that the effect of compatibility was highly attenuated when the target was unattractive. These results demonstrate that both factors were important in subjects' decisions.

Regions that selectively responded to attractiveness support our prediction that attractiveness functions like expected financial gain. Bilateral NAcc and bilateral mOFC all responded preferentially to highly attractive faces. Activation of the OFC in particular is consistent with previous studies examining activity in response to the presentation of attractive faces (O'Doherty *et al.*, 2003; Winston *et al.*, 2007), although some studies have also found NAcc activity related to attractiveness (Aharon *et al.*, 2001). This finding is consistent with a large literature indicating a 'common currency' of reward, be it social, financial, gustatory, or otherwise (Aharon *et al.*, 2001; Fehr and Camerer, 2007; Izuma *et al.*, 2008; Saxe and Haushofer, 2008; Zink *et al.*, 2008).

Activity in the right anterior insula, however, showed limited sensitivity to prospective loss information (compatibility) only when both face and compatibility information were available during Interval 2. In addition, the direction of the insular response contradicted our hypothesis. Although its involvement in differentiating compatible and incompatible information and subsequent choice supports our prediction that the insula serves similar functions in both social and financial contexts (cf. Kuhnen and Knutson, 2005), its direction did not support our hypothesis that incompatibility is a social cost.

Instead, the insula's activity in this study could be attributable to several factors. Anatomically defined ROIs necessarily introduce more noise, and the anterior insula's well known functional heterogeneity could exacerbate this problem, perhaps explaining why activity was not in the predicted direction. However, introducing noise would tend to create a bias in favor of the null hypothesis, while we observe marginally statistically significant evidence ($P < 0.055$), and the inclusion of unnecessary or irrelevant voxels was minimized by using insular sub-parcellations specific to each subject's anatomy. Alternatively, these results could be explained by subjects' interpretation of the compatibility rating. Because subjects were told explicitly that the compatibility rating was a 'good indication' of the targets' choices, compatibility could be operationalized as a potential benefit instead of a cost. In this interpretation, subjects could be construing high compatibility as an indication of a better relationship. It is also possible that just being compatible with another is intrinsically rewarding. Supporting this notion, bilateral activation of the caudate is observed in secondary whole brain analyses examining compatibility effects. The caudate has been proposed as a candidate for the processing of social reward specifically, and activity here could be related to this function (Lee, 2008). Thus, this activation lends support to the idea that compatibility could have rewarding dimensions. If insular activation does indeed represent reward-related processing, it would suggest that different types of reward processing (e.g. social and financial) are spatially dissociable, which may speak against the concept of a 'common currency' of reward processing in the ventral striatum. Yet, these interpretations are *post hoc*, and future studies should further explore the localization of different forms of reward processing and further test whether social cost can be viewed as analogous to financial loss.

The a priori regions of the bilateral NAcc and right anterior insula both predicted choice. Indeed, activity in these regions improved a logistic regression model's predictions of choice compared with using attractiveness and compatibility information alone. Thus, the predictive value of these regions cannot be explained as a simple statistical consequence of their association with attractiveness or compatibility; instead, these results indicate that activity in these regions also reflects choice-related processing in this task.

However, we acknowledge a limitation with regard to the logistic regression results. Use of individually tailored ratings of attractiveness would be preferred to the aggregate measures used the logistic regression analysis. Nevertheless, it seems unlikely that tailoring attractiveness measures would bridge the considerable difference in fit between models: the literature on facial attractiveness indicates that it is a highly reliable across cultures and individuals (Rhodes, 2006) with meta-analytic estimates of effective inter-rater reliability between 0.88 and 0.94 (Langlois *et al.*, 2000). Furthermore, using binary *vs* continuous measures of compatibility and attractiveness (which should account for more variance) did not affect discrimination between models. Finally, this issue would not affect our contention regarding the function of the insula in predicting choice, as compatibility values could not be self-generated in this study.

These regions' relationship with behavior was corroborated by a whole brain contrast that further revealed a large and robust network of frontal and subcortical regions whose activity is associated with choice in this task. This network is most clearly observed during Interval 2 Yes > No contrast, which is crucial in the exploration of the neural antecedents of social decisions, as it reflects the neural activity in the moments after subjects have received all the information they need to make their decision. Virtually all of the regions in this network have been implicated in decision-making tasks before, but in particular the regions lying in the mesolimbic dopamine pathway such as the striatum, midbrain, OFC and ACC have been frequently identified as important regions in forming choices (Montague and Berns, 2002; Sanfey, 2007; Lee, 2008; Rilling *et al.*, 2008; Rilling and Sanfey, 2011; Vickery *et al.*, 2011). Notably, this robust network is primarily observed during examination of the Interval 2 data, not Interval 1, which again serves to reaffirm the importance of the compatibility ratings in subjects' decisions—if they were ignoring these ratings, we would expect to observe similarly widespread and robust activation from crucial decision regions in the face first (Interval 1) data alone, but we do not.

A limitation of this study was that we were not able to distinguish between the immediate reward that could be conferred by viewing an attractive face (Aharon *et al.*, 2001) and the anticipated gain implied by the decision-making context. Given the number of studies that highlight the NAcc in both the outcome and anticipation phases of decision-making (Knutson *et al.*, 2001; Haber and Knutson, 2010), in addition to the impact of the immediate psychological context on reward processing (Elliott *et al.*, 2000), it is likely that the observed activity reflects both immediate and anticipated gain. This is further supported by the ventral striatum's well-established role in the processing of reward for development and execution of goal directed behavior (Haber and Knutson, 2010).

In summary, our results demonstrate that the same regions that predict choice in financial decision-making also do so in a social decision-making task in which specific choices are devoid of monetary value, although the role of the insula may be different in social tasks. Moreover, similar neural activity corresponding to anticipated gain was observed. Although both the anterior insula and NAcc, but in particular the NAcc, have been shown to be important for social decision-making, the majority of previous investigations focused on

outcome-related activity; the contribution of this study was to examine antecedent activity related to social decisions.

SUPPLEMENTARY DATA

Supplementary data are available at SCAN online.

REFERENCES

- Aharon, I., Etcoff, N., Ariely, D., Chabris, C.F., O'Connor, E., Breiter, H.C. (2001). Beautiful faces have variable reward value: fMRI and behavioral evidence. *Neuron*, 32, 537–51.
- Brainard, D.H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 443–6.
- Cho, S., Moody, T.D., Fernandino, L., et al. (2010). Common and dissociable prefrontal loci associated with component mechanisms of analogical reasoning. *Cerebral Cortex*, 20, 524–33.
- Destrieux, C., Fischl, B., Dale, A., Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage*, 53, 1–15.
- Elliott, R., Friston, K.J., Dolan, R.J. (2000). Dissociable neural responses in human reward systems. *Journal of Neuroscience*, 20, 6159–65.
- Fehr, E., Camerer, C. (2007). Social neuroeconomics: the neural circuitry of social preferences. *Trends in Cognitive Sciences*, 11, 419–27.
- Fischl, B., van der Kouwe, A., Destrieux, C., et al. (2004). Automatically parcellating the human cerebral cortex. *Cerebral Cortex*, 14, 11–22.
- Haber, S.N., Knutson, B. (2010). The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacology Reviews*, 35, 4–26.
- Izuma, K., Saito, D.N., Sadato, N. (2008). Processing of social and monetary rewards in the human striatum. *Neuron*, 58, 284–94.
- Kable, J.W., Glimcher, P.W. (2007). The neural correlates of subjective value during intertemporal choice. *Nature Neuroscience*, 10, 1625–33.
- Knutson, B., Adams, C.M., Fong, G.W., Hommer, D. (2001). Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *The Journal of Neuroscience*, 21, RC159(1–5).
- Knutson, B., Bossaerts, P. (2007). Neural antecedents of financial decisions. *The Journal of Neuroscience*, 27, 8174–7.
- Knutson, B., Rick, S., Wimmer, G.E., Prelec, D., Loewenstein, G. (2007). Neural predictors of purchases. *Neuron*, 53, 147–56.
- Kuhnen, C.M., Knutson, B. (2005). The neural basis of financial risk taking. *Neuron*, 47, 763–70.
- Langlois, J., Kalakanis, L., Rubenstein, A.J., Larson, A., Hallam, M., Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin*, 126, 390–423.
- Lebreton, M., Jorge, S., Michel, V., Thirion, B., Pessiglione, M. (2009). An automatic valuation system in the human brain: evidence from functional neuroimaging. *Neuron*, 64, 431–9.
- Lee, D. (2008). Game theory and neural basis of social decision-making. *Nature Neuroscience*, 11, 404–9.
- Levy, I., Lazzaro, S.C., Rutledge, R.B., Glimcher, P.W. (2011). Choice from non-choice: predicting consumer preferences from blood oxygenation level-dependent signals obtained during passive viewing. *The Journal of Neuroscience*, 31, 118–25.
- Manes, F., Sahakian, B., Clark, L., et al. (2002). Decision-making processes following damage to the prefrontal cortex. *Brain*, 125, 624–39.
- Minar, M., Park, D.C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers*, 36, 630–3.
- Mohr, P.N.C., Biele, G., Heekeren, H.R. (2010). Neural processing of risk. *The Journal of Neuroscience*, 30, 6613–9.
- Montague, P.R., Berns, G.S. (2002). Neural economics and the biological substrates of valuation. *Neuron*, 36, 264–84.
- O'Doherty, J., Winston, J., Critchley, H., Perrett, D., Burt, D.M., Dolan, R.J. (2003). Beauty in a smile: the role of medial orbitofrontal cortex in facial attractiveness. *Neuropsychologia*, 41, 147–55.
- Paulus, M.P., Rogalsky, C., Simmons, A., Feinstein, J.S., Stein, M.B. (2003). Increased activation in the right insula during risk-taking decision-making is related to harm avoidance and neuroticism. *Neuroimage*, 19, 1439–48.
- Paulus, M.P., Stein, M.B. (2006). An insular view of anxiety. *Biological Psychiatry*, 60, 383–7.
- Pelli, D.G. (1997). The videotoolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, 10, 437–42.
- Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annual Review of Psychology*, 57, 199–226.
- Rilling, J.K., Goldsmith, D.R., Glenn, A.L., et al. (2008). The neural correlates of the affective response to unreciprocated cooperation. *Neuropsychologia*, 46, 1256–66.
- Rilling, J.K., Sanfey, A.G. (2011). The neuroscience of social decision-making. *Annual Review of Psychology*, 62, 23–48.
- Samanez-Larkin, G.R., Hollon, N.G., Carstensen, L.L., Knutson, B. (2008). Individual differences in insular sensitivity during loss anticipation predict avoidance learning. *Psychological Science*, 19, 320–3.
- Sanfey, A. (2007). Social decision-making: insights from game theory and neuroscience. *Science*, 318, 598–602.
- Saxe, R., Haushofer, J. (2008). For love or money: a common neural currency for social and monetary reward. *Neuron*, 58, 97–8.
- Tottenham, N., Tanaka, J., Leon, A.C., et al. (2009). The NimStim set of facial expressions: judgments from untrained research participants. *Psychological Research*, 168, 242–9.
- Vickery, T.J., Chun, M.M., Lee, D. (2011). Ubiquity and specificity of reinforcement signals throughout the human brain. *Neuron*, 42, 166–77.
- Winston, J., O'Doherty, J., Kilner, J., Perrett, D., Dolan, R.J. (2007). Brain systems for assessing facial attractiveness. *Neuropsychologia*, 45, 195–206.
- Worsley, K.J., Evans, A.C., Marrett, S., Neelin, P. (1992). A three-dimensional statistical analysis for CBF activation studies in human brain. *Journal of Cerebral Blood Flow and Metabolism*, 12, 900–18.
- Zink, C.F., Tong, Y., Chen, Q., Bassett, D.S., Stein, J.L., Meyer-Lindenberg, A. (2008). Neural processing of social hierarchy in humans. *Neuron*, 58, 273–83.