Unforgettable ultimatums? Expectation violations promote enhanced social memory following economic bargaining

Luke J. Chang and Alan G. Sanfey*

Department of Psychology, University of Arizona, Tucson, AZ, USA

Recent work in the field of neuroeconomics has examined how people make decisions in interactive settings. However, less is currently known about how these social decisions influence subsequent memory for these interactions. We investigated this question by using functional magnetic resonance imaging to scan participants as they viewed photographs of people they had either recently played an Ultimatum Game with in the role of Responder, or that they had never seen before. Based on previous work that has investigated "cheater detection", we were interested in whether participants demonstrated a relative enhanced memory for partners that made either fair or unfair proposals. We found no evidence, either behaviorally or neurally, supporting enhanced memory based on the amount of money offered by the Proposer. However, we did find that participants' initial expectations about the offers they would experience in the game influenced their memory. Participants demonstrated relatively enhanced subjective memory for partners that made proposals that were contradictory to their initial expectations. In addition, we observed two distinct brain systems that were associated with partners that either offered more or less than the participants' expectations. Viewing pictures of partners that exceeded initial expectations was associated with the bilateral anterior insula, anterior cingulate cortex/premotor area, striatum, and bilateral posterior hippocampi, while viewing partners that offered less than initial expectations was associated with bilateral temporal-parietal junction, right STS, bilateral posterior insula, and precuneus. These results suggest that memory for social interaction may not be guided by a specific cheater detection system, but rather a more general expectation violation system.

Keywords: neuroeconomics, memory, social, expectation, ultimatum game, cheater detection, neural, decision-making

INTRODUCTION

Despite its relative youth, neuroeconomics as a field has made significant progress in describing the neural mechanisms that underlie decision-making (Glimcher et al., 2009). One approach within this domain has focused on circumstances in which the participant must consider the desires and intentions of another agent in reaching his or her eventual decision (Sanfey, 2007). These interactive situations have examined decisions made in a social environment, such as whether to trust or not trust another player or how to negotiate the division of a sum of money with another. The simplicity of these tasks and their ease of quantification provide not only a useful framework for developing mathematical models of optimal behavior within a social interaction (von Neumann and Morgenstern, 1944; Rabin, 1993; Fehr and Schmidt, 1999; Battigalli and Dufwenberg, 2007), but also a controlled environment within which to understand how social interaction interacts with more general cognitive processes such as memory.

One commonly used task in this domain is the Ultimatum Game (Guth et al., 1982). In this simplified bargaining scenario, one player known as the Proposer is endowed with a sum of money and told that their task is to make a proposal to the other player, the Responder, as to how this money should be divided between the two. The Proposer can make any offer he or she wants, from

keeping all of the money for themselves to giving all of it away, and any division in between. Once the offer is made, the Responder must decide to either accept or reject the proposal. If the offer is accepted, then the money is simply divided as suggested. However, if the offer is rejected, then neither player receives any money. Both players are fully aware of the rules of the game, and once the Responder makes a decision the game is over.

Many studies across a multitude of disciplines and utilizing a variety of methods have examined social decision-making using the Ultimatum Game, and the behavioral results are generally strikingly similar (Camerer, 2003). Contrary to classical predictions, which suggest that Responders should accept any non-zero offer and as a consequence Proposers should make the lowest offer possible, the modal offer to Responders is typically a little less than half of the total pot, and this amount is almost always accepted. Offers of around 30% of the pot are accepted only about half of the time, and acceptance rates diminish as offers get lower.

One suggested mechanism as to why responders turn down what is in effect 'free' money when rejecting low offers is that people severely dislike inequity (Fehr and Schmidt, 1999), and consequently feel anger in response to unfair offers (Pillutla and Murnighan, 1996; Xiao and Houser, 2005). There is compelling physiological evidence supporting this argument. Unfair offers

are associated with increased autonomic tone (van 't Wout et al., 2006) and increased activity in the anterior insula (Sanfey et al., 2003). In fact, greater insula activity in response to an unfair offer results in an increased likelihood of rejection of that offer (Sanfey et al., 2003). Other studies have found that when neural systems involved in emotion regulation are disrupted in various ways, from using tryptophan depletion (Crockett et al., 2008) to lesions of the ventromedial prefrontal cortex (Koenigs and Tranel, 2007), the result is increased rejection rates of unfair offers.

Though the response to fair and unfair offers provides an interesting window into how the competing motivations of maintaining one's reputation and maximizing one's financial gain interact in decision-making, other relevant questions can be answered using these type of tasks. Of perhaps equal importance to examining the processes that underlie performance in this task is to ask what happens, both behaviorally and neurally, when we re-encounter a player who has made a fair or unfair offer to us in the past. How do our perceptions of others shift when these people have previously treated us either well or poorly? In this initial attempt to investigate this question, we focus on memory for players with whom we have recently interacted, and specifically examine whether the way in which another player has treated us has an impact on how we in turn remember them.

Several theoretical proposals have been made as to whether we are more attuned to remembering those who have treated us either fairly or unfairly in the past. In their highly influential theory of social exchange, Cosmides and Tooby (1992) argue that humans have evolved specific cognitive abilities to promote reciprocal altruism, a construct that has been associated with positive evolutionary fitness (Trivers, 1971). Of particular importance to their theory is the ability to detect, remember, and punish "cheaters" – individuals who benefit themselves by violating a social contract (Cosmides and Tooby, 1992). However, despite the intuitive appeal of this theory, the primary evidence presented in favor of the selective detection of cheaters is that experimental participants demonstrate improved conditional reasoning when asked to detect violations of a social contract, when compared to non-social contract violations (Cosmides, 1989; Gigerenzer and Hug, 1992). Evidence supporting these theoretical claims in the domain of memory is more mixed.

Several studies have directly examined explicit memory for cheaters. There is some evidence that after 1 week participants had better memories for pictures of people with behaviors associated with cheating (e.g. "E.A. is a bishop who was caught embezzling money from his own church.") as compared to pictures of those that were associated with trustworthy behaviors (e.g. "J.H. is a vendor at baseball games who, after finding a wallet containing $250, located the owner using the driver's license.") (Mealy et al., 1996; Chiappe et al., 2004). However, more recent studies that have attempted to address some of the methodological limitations of these experiments have failed to replicate this finding, with no differences between cheaters and trustworthy pictures emerging (Barclay and Lalumiere, 2006; Mehl and Buchner, 2008). There is even some preliminary evidence for increased confidence, though not accuracy, in remembering altruists (Barclay and Lalumiere, 2006), and also that people may have better source memory than recognition memory for cheaters, meaning that people were better

at remembering that an individual was a cheater than actually correctly identifying that they had seen the person before (Buchner et al., 2009).

One explanation of these mixed findings is that the memory manipulations used were not particularly socially relevant for the participants. As outlined above, these paradigms typically involve participants reading a vignette describing either a cheating or trustworthy act by a pictured person, and then subsequently performing a recognition memory test on the set of photographs. There are surprisingly few studies that have attempted to have participants first actually engage in meaningful social interactions with other people, and then test their memory for these partners. In one study with a variant of this methodology, participants were asked to imagine playing a constant strategy (i.e. cooperate or defect) in a Prisoner's Dilemma game, and were then shown pictures and the strategies of their partners (Oda, 1997). After being tested 1 week later, the experimenters found that participants remembered defectors better than cooperators and that this effect interacted with gender. However, there was no clear explanation of the interaction with gender, nor was it clear that participants were actually engaged in the game as they were forced to stick with the same strategy.

Within neuroeconomics, there is clear evidence that people use information about a partner's history to inform decisions in future social interactions, such as to avoid trusting a cheater in a subsequent interaction or to punish them if given the opportunity. People are more likely to invest trust in partners perceived to be initially trustworthy as opposed to untrustworthy (Delgado et al., 2005; van 't Wout and Sanfey, 2008), and also seem able to then disregard this prior information when these partners actually abuse their trust. There is also evidence supporting the notion that people are willing to punish cheaters, even at the risk of incurring a financial cost to themselves in Ultimatum (Guth et al., 1982) and Public Goods Games (Fehr and Gachter, 2002; de Quervain et al., 2004). While these findings suggest that people can learn both who to trust and who not to trust and will punish cheaters given the opportunity, there is as yet no conclusive evidence directly supporting better explicit memory for either cheaters or cooperators.

One study (Singer et al., 2004) attempted to investigate this question both behaviorally and neurally by scanning participants using functional magnetic resonance imaging (fMRI) as they viewed faces which had previously behaved in either cooperative or non-cooperative ways in a modified repeated Prisoner's Dilemma game. Behaviorally, the authors report that cooperators were rated as more likeable and defectors as less likeable than control faces. In addition, participants were more accurate in recalling the behavior of both cooperators and defectors as compared to the null games. However, because there was neither money at stake for the null games nor an equal distribution of trials for each condition, the results of this forced choice memory task should be interpreted cautiously. In terms of neural findings, the authors reported that when asked to make a gender assessment of pictures of cooperators as compared to those who played null games, participants had increased activity in the left ventral putamen and left amygdala. In contrast, when participants viewed pictures of defectors compared to null trials, they showed increased activity in the vmPFC. These preliminary findings suggest that viewing faces of defectors and

cooperators from a socially relevant task may be associated with distinct neural systems. However, it remains an open question as to whether or not there may be selectively better explicit memory for cheaters and what processes might underlie this.

A possible mechanism that could explain the aforementioned pattern of results is the notion of deviation from expectation, that is, when partners play in a way differently than we predict. While it is known that alterations of expectation can affect decision-making in the Ultimatum Game (Sanfey, 2009), up to now there has been relatively little investigation of how expectations, and specifically deviation from expectations, can alter patterns of memory in social decision-making.

Some limited evidence comes from a recent study using the Trust Game, which found that participants did not have selective memory for either cooperators or defectors *per se*, but rather demonstrated enhanced memory for both types of opponents in certain circumstances, these circumstances being that the better–remembered opponent played a relatively infrequent strategy. That is, at different times they remembered both cheaters and defectors better, but only when they comprised merely 20% of the total number of interactions (Barclay, 2008). It is important to note that participants in this experiment knew *a priori* that they were playing with computer partners, so it is not clear if these results could be generalized to games played with real opponents. Nonetheless, this study provides compelling evidence that people may have enhanced memory for partners that behave contrary to social conventions, regardless of their behavior. This suggests therefore that people may not rely on a specific cheater detection system, but rather a more general expectation violation system – a notion within the field of memory that has been known for some time (von Restorff, 1933; Ranganath and Rainer, 2003), often discussed in this literature as a "novelty detection" mechanism. It is therefore possible that a more general novelty detection system can potentially be employed as a cheater detection system. Because interactions with cheaters in the real world are likely to be relatively infrequent, the expectation of cheating behavior should be low and as a result incidences of cheating should be particularly memorable. However, importantly, if we do expect substantial cheating behavior in our environment, this account would predict that partners who treat us well should be preferentially encoded and remembered.

We sought to investigate this question by using fMRI to scan the brains of participants immediately after they played a series of Ultimatum Games with a variety of partners. Firstly, we examined if our participants demonstrated more accurate memories for partners that had treated them either fairly (an equal offer) or unfairly (an unequal offer in the partner's favor). Secondly, we were particularly interested in the neural response to viewing a photograph of a previous partner as compared to a photograph of a previously unseen person, and whether the offer that had been made to the participant mediated this neural activity in our participants. Contrary to most prior behavioral studies of memory for cheaters, players in this study engaged in an actual social decision interaction and we directly assessed their social memory while they were being scanned using a standard recognition task. This study therefore can potentially inform an ongoing debate about whether people actually have enhanced memory for cheaters, and if so, whether the brain is equipped with a system to complete this task.

MATERIALS AND METHODS

PARTICIPANTS

Eighteen participants (mean age = 19.9, female = 56%) were recruited via advertisements posted on the campus of the University of Arizona to participate in this study. All participants were screened for any significant health-related or neuropsychiatric disorders and none were currently taking psychoactive medication. Two participants were excluded from the analysis for technical reasons (corrupted data). All participants gave informed consent according to procedures approved by the University of Arizona's Institutional Review Board.

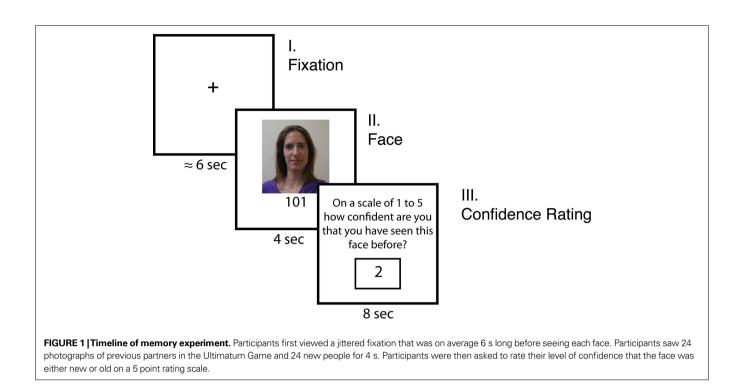### PROCEDURE

Expectations

Prior to being scanned, we elicited participants' beliefs about the kinds of offers they expected to encounter, with participants being asked the number of people out of 100 that they believed would make a $0, $1, $2, $3, $4, $5, $6, or $7 offer. Participants' elicited expectations prior to playing the game were used to create a distribution of the frequency of offers that they expected to encounter. The mode of this distribution was used to represent each participant's initial expectation.

Ultimatum game

Participants then played a standard single-shot Ultimatum Game in the role of Responder with 48 different partners while undergoing fMRI. Twenty-four of these partners were human, 12 were computers, and 12 were non-intentional humans (i.e. humans whose responses were randomly generated). Each offer was preceded by a picture of their partner for that round. Though participants were told that the human-intentional offers would be made by other players, in actual fact all offers were controlled by the experimenter, and all participants saw the same set of offers. This set consisted of equal numbers (12 each) of $1, $2, $3 and $5 offers, all of which were made from a $10 pot. For each participant, all pictures were randomly paired to an offer amount, ensuring that there was no potential picture by offer amount interaction. Participants were paid $20 for participating and an additional $5, which they believed was based on their performance in the game. Further details of the Ultimatum Game portion of the experiment will be described in greater detail in a separate paper. While participants were not directly queried after the experiment about whether or not they believed that they were interacting with real partners, no participant expressed doubt towards the experimenter at any time during the experiment.

Memory experiment

After completing the Ultimatum Game trials, participants were given an incidental memory test while undergoing fMRI (see **Figure 1** for a trial timeline). Participants had not been forewarned about this task. Participants viewed randomly presented pictures of the 24 human-intentional partners they had previously encountered, as well as 24 new faces they had not seen in the Ultimatum Game task. Participants did not view pictures of the computers or the non-intentional human partners. There were an equal number of male and female faces for both the new and old sets of faces. To begin each trial, a jittered fixation was seen for 6 s on average

FIGURE 1 | Timeline of memory experiment. Participants first viewed a jittered fixation that was on average 6 s long before seeing each face. Participants saw 24 photographs of previous partners in the Ultimatum Game and 24 new people for 4 s. Participants were then asked to rate their level of confidence that the face was either new or old on a 5 point rating scale.

and then a face was shown for 4 s. After being presented with this photograph, participants were asked to rate their confidence that they had played the Ultimatum Game with this person on a scale of 1 to 5 (1 = "I've definitely never seen this person before"; 2 = "I don't think I've seen them before"; 3 = "I'm not sure if I've seen this person before"; 4 = "I think I may have seen them before"; 5 = "I've definitely seen this person before"). Participants were given 8 s to make this judgment, and ratings were entered by scrolling through the possible options with one response button and then selecting the desired rating with the other button. While the rating system was always the same (e.g. 1 = never seen; 5 = definitely seen), the ratings randomly scrolled up or down on each trial to eliminate any possible motor confounds. Therefore, the number of button presses were orthogonal to the actual ratings provided. Stimuli were presented via EPrime software (Psychology Software Tools, Inc, Pttsburg, PA, USA) using MRI-compatible googles and responses were recorded using a fiber optic button box (Resonance Technologies, Van Nuys, CA, USA).

ANALYSES

All behavioral statistics were computed using the R statistical package (R_Development_Core_Team, 2008). For regressions that included repeated observations, we used the lme4 mixed effects general linear model package (Bates et al., 2008). Participants were treated as a random effect with varying intercepts and slopes. We report the parameter estimates (b), standard error, t-values, and p-values. Because there is no generally agreed upon method for calculating p-values in mixed models, we used two separate methods. First, we calculated the degrees of freedom by subtracting the number of observations minus the number of fixed effects (Kliegl et al., 2007). Second, we generated confidence intervals from the posterior distribution of the parameter estimates using

Markov Chain Monte Carlo methods (Baayen et al., 2008). These results were identical unless otherwise noted. For robust regressions we used the rlm function from the MASS package using an MM-estimator (Venables and Ripley, 2002).

D′

To measure participant's ability to discriminate old from new faces, we used D′, a signal detection metric (Wickens, 2002). D′ controls for individual participants' response bias (i.e. their propensity to say yes) and was calculated as the difference between the standardized z-score for hits (indicated by a 4 or 5 on the confidence rating for an old face) and the standardized z-score for false positives (indicated by a 4 or 5 on the confidence rating for a new face). Because this analysis emphasizes hits and false positives, it ignores differences in levels of subjective confidence, that is, the difference between a 1 and 2, or a 4 or 5 rating. D′ scores were calculated separately for every level of offer amount.

Data acquisition

Each scanning session included a T1-weighted MPRAGE structural scan (TR = 11 ms, TE = 4 ms, matrix = 256 × 256, slice thickness = 1 mm, gap = 0 mm), followed by five functional runs. The first 3 functional runs contained the Ultimatum Game trials and the last two contained the memory trials (240 volumes per run). Functional scans used a 3-shot multiple echo planar imaging (MEPI) GRAPPA sequence using parameters selected to maximize signal in regions associated with high susceptibility artifact, such as orbitofrontal cortex and medial temporal lobe (Stocker et al., 2006; Weiskopf et al., 2006) (TR = 2000 ms, TE = 256 ms, matrix = 96 × 96, FOV = 192 mm, slice thickness = 3.0 mm, 42 axial slices, voxel size 2 × 2 × 3). The MEPI sequence employs parallel imaging and allows for increases in

signal intensity, image resolution, the number of slices that can be acquired in a 2000-ms TR, as well as substantial decreases in geometric distortion (Newbould et al., 2007).

Data preprocessing

Functional imaging data were preprocessed and analyzed using the FSL Software package 4.1.4 (FMRIB, Oxford, UK). The first three volumes of the functional runs were discarded to account for T1 equilibrium effects. Images were corrected for slice scan time using an ascending interleaved procedure. Head motion was corrected using MCFLIRT using a 6-parameter rigid-body transformation. Images were spatially smoothed using a 5 mm full width at half maximum Gaussian kernel. A high pass filter was used to cut off temporal periods longer than 66 s. All images were initially co-registered to the participant's high resolution structural scan and were then co-registered to the MNI 152 person 2-mm template using a 12-parameter affine transformation. All functional analyses are overlaid on the participants' average high resolution structural scan in MNI space.

General imaging analysis methods

A 3-level mixed effects general linear model (GLM) was used to analyze the imaging data. A first-level GLM was defined for each participant's functional run that included a boxcar regressor for each epoch of interest (e.g. face phase), convolved with a canonical double-gamma hemodynamic response function. To account for residual variance we also included the temporal derivatives of each regressor of interest, the six estimated head movement parameters, and any missed trials as covariates of no interest. The resulting GLM was corrected for temporal autocorrelations using FILM prewhitening. A second-level fixed effects model was fit for each participant to account for intra-run variability. For each participant, contrasts were calculated between predictors for different regressors of interest at every voxel in the brain. A one-sample t-test was used at the third-level for each contrast using a Bayesian implementation of mixed effects inference (Behrens et al., 2008). We corrected for multiple comparisons with cluster correction utilizing Gaussian random field theory with an initial cutoff of $Z > 2.3$ and a FWE $p < 0.05$.

We report the results of three analyses. The offer amount analysis included individual regressors during the face phase for players who had previously made offers of $1, $2, $3, or $5, a regressor indicating the duration of the response time during the memory phase, a regressor indicating a distractor face, a regressor for missed trials, the temporal derivatives of each of these predictors and 6 motion parameters (20 predictors total). We report the results for the Unfair (i.e. $1 and $2 offers) vs Fair (i.e. $5) contrast. For the expectation violation analysis we included regressors at the face phase for offers below expectation (i.e. standardized expectation error (SEE) > 0), offers above expectation (i.e. SEE < 0), and offers at expectation (i.e. SEE = 0). SEE is the within-subject z-score of the numerical deviation of an offer amount from a participant's initial expectations. In addition, we included a regressor modeling the memory phase for the duration of the response, a regressor indicating a distractor face, a regressor modeling missed trials, and their temporal derivatives and 6 motion parameters (18 predictors total). We report a linear contrast of prediction error (i.e. +1 0 −1) for Positive, Zero, and Negative SEE regressors. Finally, the third

analysis was identical to the second analysis except the third level linear contrast was weighted by each participant's standardized initial expectation, effectively utilizing a correlation analysis rather than a one sample t-test. This analysis tests the interaction between participant's initial expectation and their SEE. All trials in which the participants indicated that they were unsure (i.e. a rating of 3) were excluded from all analyses (78 trials total for all subjects, or 10.2% of observations).

RESULTS
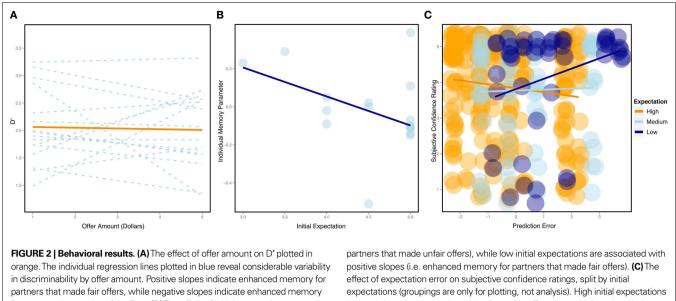
BEHAVIORAL RESULTS

Ultimatum game

Consistent with previous research (Sanfey et al., 2003; van 't Wout et al., 2006; Harle and Sanfey, 2007), acceptance rates decreased as offers got lower, and participants were significantly more likely to accept fair ($5) as opposed to unfair ($1; $2) offers, illustrated using a mixed effects logit model (Jaeger, 2008), b = 4.24, se = 0.84, odds ratio = 69.21, Wald Z = 5.07, $p < 0.05$. The average acceptance rate for all intentional offers was 62.2%, with three participants accepting all offers. Consistent with previous research, participants expected most participants to make fair offers (mean = 4.5, sd = 0.63) (Sanfey, 2009).

Memory

A one-sample t-test revealed that participants were on average accurate in their ability to discriminate between old and new faces (mean D′ = 2.04), $t(15)$ = 14.68, $p < 0.001$. Participants were able to correctly identify both old faces (mean correct = 70%, se = 0.04) and new faces (mean correct = 76%, se = 0.04). We used a mixed effects linear model to test whether or not the amount of money offered by the proposer would influence participant's memory for that person, but did not observe a significant effect, b = 0.03, se = 0.05, t = 0.52, ns. These results indicate that participants were sensitive in their ability to discriminate between new and old faces, but that, on average, this discriminability was not influenced by the amount of money offered by the partner.

However, closer examination of these results indicate that participants demonstrated considerable variability in their ability to remember proposers that made either fair or unfair offers (**Figure 2**). Some participants appeared to demonstrate improved memory for proposers that made unfair offers, while other participants remembered proposers that made fair offers indicated by the random effect slope coefficient). Using robust regression we found that participants' initial expectations predicted the random effects parameter estimates from the previous offer amount analysis, parameter estimate = 0.16, se = 0.04, t = 4.32, $p < 0.05$. This analysis indicates that as initial expectations increased, the slope of offer amount on D′ decreased. In other words, participants with low initial expectations had positive memory slopes, meaning that they demonstrated augmented memory for proposers that made fair offers, while participants with high initial expectations had negative memory slopes, indicating increased memory for proposers that made unfair offers (**Figure 2A**).

To test this expectation violation hypothesis more explicitly, we used a mixed effects linear model treating subjects as a random intercept. Specifically, we attempted to predict participant's

FIGURE 2 | Behavioral results. (A) The effect of offer amount on D′ plotted in orange. The individual regression lines plotted in blue reveal considerable variability in discriminability by offer amount. Positive slopes indicate enhanced memory for partners that made fair offers, while negative slopes indicate enhanced memory for partners that made unfair offers. **(B)** The effect of initial expectation on individual memory parameter [i.e. the slopes from **(A)**]. The individual variability in memory for offer amounts is related to initial expectations. High initial expectations are associated with negative slopes (i.e. enhanced memory for partners that made unfair offers), while low initial expectations are associated with positive slopes (i.e. enhanced memory for partners that made fair offers). **(C)** The effect of expectation error on subjective confidence ratings, split by initial expectations (groupings are only for plotting, not analysis). High initial expectations (orange) demonstrated enhanced memory for offers that were less than their initial expectation. Low initial expectations (dark blue) were associated with enhanced memory for partners that made offers that exceeded their expectations. No relationship was found for medium initial expectations (light blue).

subjective confidence ratings using their centered initial expectation, the centered deviation of the offer amount from their initial expectation, and the interaction between these two variables. We observed an initial expectation by expectation deviation interaction, $b = 0.17$, se $= 0.08$, $t = 2.21$, $p < 0.05$, with no significant main effects. Participants demonstrated enhanced confidence ratings for faces that violated their initial expectations (see **Figure 2**).

IMAGING RESULTS
Offer amount
As noted above, we observed no significant effect of offer amount in predicting participant's ability to discriminate between old and new faces. Similarly, for the corresponding imaging analysis, we did not observe any significant voxels above threshold for this previously fair vs. previously unfair contrast, even at a more liberal $p < 0.001$ (uncorrected) threshold. Therefore, at least on average across participants, there is no particular neural signature for either previously fair or previously unfair partners.

Expectation violation
Our more detailed behavioral analysis indicated that expectation violation, and not offer amount, was associated with enhanced subsequent memory for partners. To explore the neural systems underlying this effect we ran two separate imaging analyses. The first analysis examined the effect of expectation deviation. This contrast was associated with bilateral anterior insula, pre-supplementary motor area (pre-SMA)/anterior cingulate cortex (ACC), the striatum (including the caudate and nucleus accumbens), and bilateral posterior hippocampi/parahippocampi. Negative expectation deviations were associated with bilateral temporal parietal junction (TPJ), right superior temporal sulcus

(STS), posterior insula, and precuneus (see **Figure 3**). The second analysis examined the interaction between the initial expectation and the expectation error, by weighting the first analysis by each participant's standardized initial expectation at the third level. No voxels survived our threshold for this analysis. Thus, this set of analyses reveals a network previously associated with expectation violation (i.e. insula, pre-SMA, and NAcc), and memory retrieval (i.e. hippocampi/parahippocampi) when participants view faces of partners who offered more than the participants initially expected, and a network associated with theory of mind processing (i.e. STS/TPJ) and memory (i.e. precuneus) when viewing partners that offered less than the participant initially expected (see **Table 1** for a complete list of regions).

DISCUSSION
This study investigated how economic exchange impacts subsequent memories for social partners. Following a standard Ultimatum Game paradigm, participants were shown photographs of both previously seen and unseen people, and asked to rate their confidence that they had viewed these pictures before. This question is important in understanding the behavioral and neural effects of reappraising a partner with whom one has previously been engaged in social economic interaction. In addition, this research was also interested in investigating the notion of "cheater detection", that is, the idea of relatively enhanced memories for social partners who have treated us badly in the past (Cosmides and Tooby, 1992; Mealy et al., 1996; Singer et al., 2004; Barclay, 2008).

We were primarily interested in whether participants exhibited a relative memory enhancement for partners that made either fair or unfair proposals. A demonstration of the latter (i.e. enhanced memory for unfair proposers) would provide evidence supporting the existence of behavioral cheater-detection effects. However, we
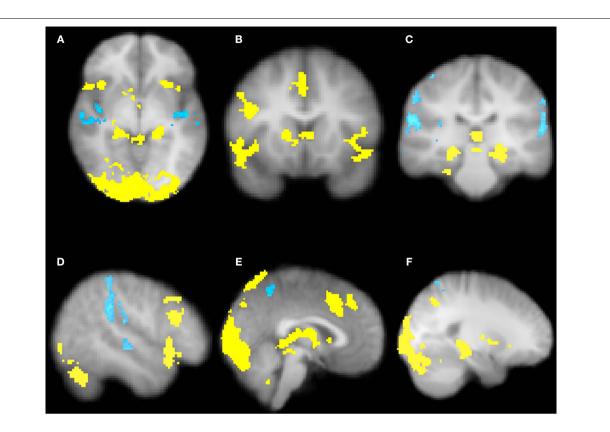
FIGURE 3 | Brain regions associated with expectation violation. These results are a linear contrast of standardized expectation error (SEE) when participants are viewing pictures of their partners during the memory task. Yellow values are associated with partners that offered more than the participant's initial expectation. Blue values are associated with viewing partners that offered less than the participant's initial expectation. **(A)** Axial section shows bilateral insula, ventral striatum, bilateral posterior hippocampi, visual cortex associated with positive SEE, while right superior temporal sulcus and bilateral posterior insula are associated with negative SEE. **(B)** Coronal section shows bilateral anterior insula, ventral striatum, right dorsolateral prefrontal cortex, and pre-supplementary motor area/dorsal anterior cingulate. **(C)** Coronal section shows bilateral posterior hippocampus. **(D)** Right lateral section shows DLPFC, VLPFC, and TPJ and STS. **(E)** Sagittal section shows preSMA/DACC. **(F)** Sagittal section shows posterior hippocampus/parahippocampus. All clusters survive whole brain correction using cluster correction, $Z > 2.3$, $p < 0.05$ and are displayed on the group average T1 image using the radiological convention (left = right).

did not observe a significant effect, either behaviorally or neurally, for an influence of offer fairness on memory. Instead, we found that participants demonstrated considerable variability in their ability to discriminate between partners associated with different levels of fairness (e.g. some participants demonstrated relative enhanced memory for fair partners and some for unfair partners). Importantly, this variability was predicted by their initial expectations about the range of offers they would see in the game. Those who had low initial expectations (operationalized by their modal reported expected offer) were more likely to demonstrate augmented memory for partners that exceeded their expectations, while those who had high initial expectations demonstrated better memory for partners who made offers lower than their initial expectations. We also observed a significant interaction between participants' initial expectations and their expectation error (i.e. the proposal's deviation from initial expectation) in predicting subjective confidence ratings. Participants who had low initial expectations were more likely to remember partners that made offers that were greater than their initial expectations, while participants that had high initial expectations were more likely to remember partners that made offers that were lower than their initial expectations.

This finding is consistent with the results of a recent cheater detection study (Barclay, 2008), in which participants demonstrated enhanced memory in a recognition test for whichever behavior (e.g. cooperate or defect) was more *infrequent* in a previously-played Trust Game. When the majority of partners cooperated, participants remembered defectors better, whereas when most partners defected, participants remembered cooperators better. Our study employed a different approach than that of (Barclay, 2008), namely use of an Ultimatum as opposed to the Trust Game, and additionally we used the participants own expectations as the "baseline", as opposed to examining violations from experienced probabilities, but the two sets of results converge on the same interpretation – that deviations from prior expectations result in greater salience and thus better memory encoding.

This conjecture has been posited for a long time in the memory literature dating to von Restorff (1933). The Von Restorff effect refers to memory enhancement occurring when an item is isolated either by manipulating the context (e.g. item is printed in red in a list of items printed in black) or content (e.g. inserting a nonsense syllable into a list of meaningful words) (Wallace, 1965). This effect is thought to be associated with unexpected change rather than

Table 1 | Brain regions associated with expectation error.

Hemisphere	Region	BA	*Z* value	X	Y	Z
POS > NEG						
L	Angular gyrus	39	3.56	−46	−60	36
L	Anterior insula	48	3.15	−42	12	−4
L	Frontal operculum	48	3.04	−42	18	4
L	Lateral OFC	38	3.23	−38	18	−14
L	Midbrain (substantia nigra)	NA	3.76	−10	−14	−10
L	Occipital cortex	18	4.88	−34	−94	10
L	Occipital cortex (primary visual)	17	4.99	−6	−96	14
L	Posterior hippocampus	27	4.14	−20	−32	−4
L	Superior parietal lobule	7	3.6	−30	−58	42
L	Temporal pole	38	3.36	−50	14	−8
R	ACC	24	3.73	2	20	36
R	Anterior insula	47	3.44	42	16	−8
R	Fusiform gyrus	37	5.03	38	−46	−24
R	Inferior frontal gyrus	48	3.55	54	18	20
R	Middle frontal gyrus	44	3.04	52	14	36
R	Occipital cortex	18	5.11	8	−96	20
R	Parahippcampus	27	4.19	18	−34	−8
R	Posterior hippocampus	20	4.04	24	−26	−10
R	Pre-SMA	32	3.84	4	20	44
R	Cerebellum (right VI)	19	5.08	28	−68	−20
R	Superior frontal gyrus	8	3.31	0	36	50
R	Temporal pole	38	3.43	50	20	−20
NEG > POS						
L	Angular gyrus	39	3.56	−46	−60	36
L	Posterior insula	48	3.01	−38	−12	−2
L	Precuneus	7	3.22	−6	−60	52
L	Superior parietal lobule	5	3.17	−18	−60	66
L	Superior temporal gyrus	22	3.5	−62	−30	12
L	STS	48	3.73	−48	−12	−8
L	Supramarginal gyrus anterior division (TPJ)	40	3.39	−62	−30	40
L	TPJ (parietal operculum cortex)	48	3.41	−58	−38	26
R	Posterior insula	20	3.79	40	−12	−10
R	Precuneus	5	3.08	4	−48	60
R	Superior temporal gyrus	42	3.47	58	−34	16
R	STS	22	3.5	62	−14	−6
R	Supramarginal gyrus posterior division (TPJ)	48	3.77	54	−36	28
R	TPJ (parietal operculum cortex)	48	3.79	62	−28	22

This table reflects the contrast positive expectation error compared to negative expectation error and shows the local maxima of clusters surviving cluster correction Z > 2.3, p < 0.05 in MNI space. Cortical and subcortical regions were identified using the Harvard-Oxford Probabilistic Anatomical Atlas and Mai et al. (2007), while the cerebellar regions were identified using a probabilistic cerebellar atlas (Diedrichsen et al., 2009). Abbreviations: TPJ = temporal-parietal junction, SMA = supplementary motor area, STS = superior temporal sulcus, OFC = orbitofrontal cortex, ACC = anterior cingulate cortex.

actual isolation (Green, 1956). The source of this mechanism has been the focus of considerable research in the memory, attention, and cognitive control literatures and has even served as one of the primary paradigms in studying cognition in preverbal infants (Fantz, 1964). Detecting novel stimuli embedded within more frequent background stimuli has been extensively studied using a paradigm known as the "oddball task". The ability to detect novel stimuli or expectation violations is associated with a distinct event related potential that occurs about 300 ms after the novel stimulus

onset (Sutton et al., 1965). While the precise neural origins of this signal are still being worked out (Ranganath and Rainer, 2003), the hippocampus (Knight, 1996; Tulving et al., 1996), ACC (Baudena et al., 1995; Berns et al., 1997), and insula (Linden et al., 1999; Kiehl et al., 2001) have been shown to be reliably involved. Our findings support the existence of this more general system that detects violations of expectations and, importantly, extends these ideas into the domain of social interactive decision-making and neuroeconomics.

In terms of our imaging results, we also found distinct networks consistent with systems previously identified with expectation violation and memory. Viewing faces of partners whose offers exceeded expectations was associated with bilateral posterior hippocampi/parahippocampi, bilateral anterior insula, pre-SMA/ACC, and striatum. These regions have previously been associated with expectation violation, social cognition, and memory. Considerable research has demonstrated that posterior hippocampal regions are critical in successful recognition memory (Eldridge et al., 2000; Yonelinas et al., 2005). In addition, patients with hippocampal damage have been demonstrated to have a selective impairment in generating the characteristic P300 and autonomic skin response following unexpected events while the processing of expected events remained preserved (Knight, 1996). Our observed activation in the striatum, which included the caudate and nucleus accumbens is consistent with the literature on reward prediction error (Schultz et al., 1997) and repeated play with cooperators (Rilling et al., 2002). Partners that exceed initial expectations are associated with a positive prediction error, which is likely to promote further cooperation with these partners in the future (Rilling et al., 2002; Delgado et al., 2005; King-Casas et al., 2005). We also observed activity in the pre-SMA area/ACC and bilateral anterior insula. This network appears to be functionally coupled (Fox et al., 2005; Margulies et al., 2007; Craig, 2009), and has consistently been associated with detecting violations of expectation in a multitude of contexts including stimulus frequency (Braver et al., 2001), changes in sequences (Berns et al., 1997; Huettel et al., 2002) and multi-modal sensory changes (Downar et al., 2000). Thus, viewing pictures of partners who exceeded participants' expectations resulted in increased activity in regions of the brain that have been consistently associated with detecting violations of expectations in paradigms investigating more basic aspects of novelty detection and also in successful memory retrieval.

In contrast, when viewing partners that had made lower offers than the player had expected, we found activation in bilateral TPJ, right STS, bilateral posterior insula, and precuneus. These regions have been implicated in a variety of processes including memory, expectation violation, social cognition, and pain processing. The TPJ has been shown to be involved in expectation violation (Downar et al., 2000) and plays a key role in generating the brain's P300 novelty response (Knight et al., 1989) and in orienting attention (Corbetta et al., 2000). In addition, the TPJ has received attention for its role in thinking about others' mental states (i.e. theory of mind) (Saxe and Kanwisher, 2003), but it is currently unclear if these two processes can be explained by a more general cognitive process (Mitchell, 2008). Thus, viewing pictures of partners who offered less money than was expected is associated with a region of the brain that has been implicated in both social cognition and novelty detection. We also observed increased activity in the right STS, a region which has been hypothesized to detect and evaluate intentions and actions of other's behavior (Frith and Frith, 1999; Saxe et al., 2004). This region has been associated with updating expectations about an opponent's strategy based on their behavior in a repeated Inspection game (Hampton et al., 2008). In addition, the STS and TPJ have been demonstrated to be involved in social

prediction error – specifically in both making a prediction about the value of a social partner's advice and updating this prediction after feedback (Behrens et al., 2008). We also observed activity in the bilateral posterior insula, which has been primarily associated with interoceptive processing, that is processing of the physiological condition of the body (Craig, 2002). This region is reliably associated with processing pain from external stimulation (Koyama et al., 2005; Singer et al., 2004) and also direct cortical stimulation (Ostrowsky et al., 2002) and suggests, at least tentatively, that viewing pictures of participants who offered less than expectations is perhaps associated with processing a negative somato-visceral state. Finally, the precuneus has been demonstrated to be involved in memory, and social cognition (Cavanna and Trimble, 2006). The range of these memory processes is diverse and includes episodic memory retrieval (Shallice et al., 1994; Tulving et al., 1994), recognition memory (Henson et al., 1999; Yonelinas et al., 2005), source memory (Lundstrom et al., 2005), and autobiographical memory retrieval (Addis et al., 2004). The precuneus has also been involved in mentalizing perceived intentionality (den Ouden et al., 2005), and in reasoning about another's beliefs (Saxe and Kanwisher, 2003). These results suggest that viewing a picture of a partner that offered less money than was initially expected is associated with brain regions that have been thought to be involved in processing negative somatic states, mentalizing about another person's beliefs, updating expectations about behavior, and memory.

Interestingly, despite the methodological differences between the present study and that of Singer et al. (2004), both studies yield somewhat similar results. Singer et al. (2004) had participants repeatedly make a gender discrimination on photographs of partners with whom they had previously encountered in a repeated Prisoner's Dilemma Game. In contrast, our study employed a single-shot design using the Ultimatum Game and a recognition task that included an equal number of old and new faces. Our imaging analyses focus on partners that made offers that were either higher or lower than the participant's initial expectation, while Singer et al. (2004) independently compared partners that were cooperators or defectors to partner's associated with null games. Despite these methodological discrepancies both studies identify the anterior insula and different components of the striatum as being linked to partners with positive associations (i.e. cooperator or positive expectation error). While Singer et al. (2004) only found vmPFC associated with defectors, we observed activity in the posterior insula, STS, TPJ. Because our study was explicitly designed to study social memory, we were also able to observe activity in regions that have previously been associated with memory retrieval – most notably the hippocampal/parahippocampal regions and precuneus. Thus, our results extend those of Singer et al. (2004), by providing a different perspective on cheater detection (i.e. expectation violation) as well as methods that are more conducive to studying social memory.

In contrast to our behavioral results, we did not observe activation in the brain for the interaction between initial expectations and expectation error. One possible reason why we failed to observe a significant finding for the imaging interaction is the combination of a stringent statistical threshold and a lack of

statistical power. While our mixed effects procedure can account for unequal variances in the interaction analysis, there is an under representation of cases in which there are low initial expectations in this sample. Thus, while our behavioral analyses utilized a p-value of $p < 0.05$, our imaging analyses were restricted to a more stringent criteria to account for multiple comparisons. Indeed, when the statistical threshold is dropped to a more liberal $p < 0.005$ uncorrected level, we find almost identical results to our expectation error analysis including bilateral anterior insula, SMA, bilateral posterior hippocampus, bilateral caudate, left ventral putamen, and bilateral amygdala.

As noted, we did not observe any evidence for a significant behavioral or imaging finding for enhanced memory for partners based on the amount of money they offered. Nor did we observe evidence of a salience detection system, in which partners who made either extremely fair or unfair offers were better remembered. The literature on cheater detection is rife with conflicting results, with some studies finding enhanced memory for cheaters (Mealy et al., 1996; Oda, 1997; Chiappe et al., 2004), others finding enhanced memory for altruists (Barclay and Lalumiere, 2006), and others, like the present study, finding no significant differences (Barclay and Lalumiere, 2006; Mehl and Buchner, 2008). The more general expectation deviation system outlined here is a potential mechanism that could account for the inconsistent results in this domain. However, at present it is not immediately clear why we identified two distinct expectation violation systems that track with the valence of the deviation. In addition, it is important to note that despite the attractiveness of the expectation violation hypothesis, our imaging results do not necessarily rule out the possibility that some of the regions associated with negative expectation violations may be involved in cheater detection. Addressing these issues could be fruitfully explored further in future research.

Like all studies, there are a number of limitations that should be considered before drawing firm conclusions from the results. First, it is always difficult to interpret null findings. Our lack of significant results for offer amount on memory cannot necessarily be interpreted as an absence of an effect. Neuroimaging studies are inherently underpowered (Mumford and Nichols, 2008) and as such are greatly at risk for making Type II error. Second, it is unclear if participants actually believed they were engaged in a real social interaction. While participants were not explicitly probed about whether they believed that they were playing with a real partner, no participant expressed any doubt, nor did their behavior deviate remarkably from other published studies that utilized actual human partners (Camerer, 2003). In addition, consistent with previous research (Sanfey, 2009), most participants indicated that they expected their partners to make fair offers. Finally, it is somewhat of an open question as to whether a single UG interaction is sufficient to label a partner as a "cheater". It is possible that making such a judgment would require multiple interactions. However, a single interaction would be enough to develop an initial impression and there is considerable evidence demonstrating that participants generate negative emotional responses in response to a single unfair offer (Sanfey et al., 2003; van 't Wout et al., 2006).

In summary, our results support a more general system that detects violations of expectations as opposed to a more specialized system engineered to detect cheaters. We found that participants on average were no better or worse at remembering partners who made either fair or unfair proposals, but rather that individual participants exhibited selectively better memory for partners who made offers which violated their initial expectations. Two dissociable neural systems were found to be underlying this effect. While both systems have been previously associated with expectation violation, social cognition, and memory, these regions tentatively suggest that there is distinct processing for positive and negative expectation violations. Positive expectation violations are associated with a system that may incorporate error detection, conscious awareness of the error, reward processing, and enhanced recognition memory, while negative expectation violations are associated with expectation violation, evaluating intentions, pain processing, and autobiographical episodic memory. By incorporating the strengths of several fields – the tasks of behavioral economics, the methodologies of psychology and the sophisticated techniques of neuroscience – we can uniquely investigate how social exchange operates, not just in terms of the immediate decisions but also how these interactions can reverberate over time.

ACKNOWLEDGMENTS

REFERENCES

Addis, D. R., McIntosh, A. R., Moscovitch, M., Crawley, A. P., and McAndrews, M. P. (2004). Characterizing spatial and temporal features of autobiographical memory retrieval networks: a partial least squares approach. *Neuroimage* 23, 1460–1471.

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 340–412.

Barclay, P. (2008). Enhanced recognition of defectors depends on their rarity. *Cognition* 107, 817–828.

Barclay, P., and Lalumiere, M. L. (2006). Do people differentially remember cheaters? *Hum. Nat.* 17, 98–113.

Bates, D., Maechler, M., and Dai, B. (2008). lme4: Linear Mixed-Effects Models Using S4 Classes. R package version 0.999375-28.

Battigalli, P., and Dufwenberg, M. (2007). Guilt in games. *Am. Econ. Rev.* 97, 170–176.

Baudena, P., Halgren, E., Heit, G., and Clarke, J. M. (1995). Intracerebral potentials to rare target and distractor auditory and visual stimuli. III. Frontal cortex. *Electroencephalogr. Clin. Neurophysiol.* 94, 251–264.

Behrens, T. E., Hunt, L. T., Woolrich, M. W., and Rushworth, M. F. (2008). Associative learning of social value. *Nature* 456, 245–249.

Berns, G. S., Cohen, J. D., and Mintun, M. A. (1997). Brain regions responsive to novelty in the absence of awareness. *Science* 276, 1272–1275.

Braver, T. S., Barch, D. M., Gray, J. R., Molfese, D. L., and Snyder, A. (2001). Anterior cingulate cortex and response conflict: effects of frequency, inhibition and errors. *Cereb. Cortex* 11, 825–836.

Buchner, A., Bell, R., Mehl, B., and Musch, J. (2009). No enhanced recognition memory, but better source memory for faces of cheaters. *Evol. Hum. Behav.* 30, 212–234.

Camerer, C. F. (2003). Dictator, ultimatum, and trust games. In Behavioral Game Theory, C. F. Camerer, ed. (New York, Russell Sage Foundation), pp. 43–116.

Cavanna, A. E., and Trimble, M. R. (2006). The precuneus: a review of its functional anatomy and behavioural correlates. *Brain* 129(Pt 3), 564–583.

Chiappe, D., Brown, D., Dow, B., Koontz, J., Rodriguez, M., and McCulloch, K. (2004). Cheaters are looked at longer and remembered better than

cooperators in social exchange situations. *Evol. Psychol.* 2, 108–120.

Corbetta, M., Kincade, J. M., Ollinger, J. M., McAvoy, M. P., and Shulman, G. L. (2000). Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nat. Neurosci.* 3, 292–297.

Cosmides, L. (1989). The logic of social exchange: has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition* 31, 187–276.

Cosmides, L., and Tooby, J. (1992). Cognitive adaptations for social exchange. In The Adapted Mind: Evolutionary Psychology and the Generation of Culture, J. Barkow, L. Cosmides and J. Tooby, eds (New York, Oxfor University Press). pp. 163–228.

Craig, A. D. (2002). How do you feel? Interoception: the sense of the physiological condition of the body. *Nat. Rev. Neurosci.* 3, 655–666.

Craig, A. D. (2009). How do you feel—now? The anterior insula and human awareness. *Nat. Rev. Neurosci.* 10, 59–70.

Crockett, M. J., Clark, L., Tabibnia, G., Lieberman, M. D., and Robbins, T. W. (2008). Serotonin modulates behavioral reactions to unfairness. *Science* 320(5884), 1739.

de Quervain, D. J., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., and Fehr, E. (2004). The neural basis of altruistic punishment. *Science* 305, 1254–1258.

Delgado, M. R., Frank, R. H., and Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nat. Neurosci.* 8, 1611–1618.

den Ouden, H. E., Frith, U., Frith, C., and Blakemore, S. J. (2005). Thinking about intentions. *Neuroimage* 28, 787–796.

Diedrichsen, J., Balsters, J. H., Flavell, J., Cussans, E., and Ramnani, N. (2009). A probabilistic MR atlas of the human cerebellum. *Neuroimage* 46, 39–46.

Downar, J., Crawley, A. P., Mikulis, D. J., and Davis, K. D. (2000). A multimodal cortical network for the detection of changes in the sensory environment. *Nat. Neurosci.* 3, 277–283.

Eldridge, L. L., Knowlton, B. J., Furmanski, C. S., Bookheimer, S. Y., and Engel, S. A. (2000). Remembering episodes: a selective role for the hippocampus during retrieval. *Nat. Neurosci.* 3, 1149–1152.

Fantz, R. L. (1964). Visual Experience in Infants: Decreased Attention to Familiar Patterns Relative to Novel Ones. *Science* 146, 668–670.

Fehr, E., and Gachter, S. (2002). Altruistic punishment in humans. *Nature* 415, 137–140.

Fehr, E., and Schmidt, K. M. (1999). A Theory of Fairness, Competition, and Cooperation. Boston, MIT Press.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., and Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc. Natl. Acad. Sci. U.S.A.* 102, 9673–9678.

Frith, C. D., and Frith, U. (1999). Interacting minds–a biological basis. *Science* 286, 1692–1695.

Gigerenzer, G., and Hug, K. (1992). Domain-specific reasoning: social contracts, cheating, and perspective change. *Cognition* 43, 127–171.

Glimcher, P. W., Camerer, C., Poldrack, R. A., and Fehr, E. eds (2009). Neuroeconomics: Decision-making and the brain. New York, Academic Press. 538p.

Green, R. T. (1956). Surprise as a factor in the von restorff effect. *J. Exp. Psychol.* 52, 310–344.

Guth, W., Schmittberger, R., and Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *J. Econ. Behav. Organ.* 3, 367.

Hampton, A. N., Bossaerts, P., and O'Doherty, J. P. (2008). Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proc. Natl. Acad. Sci. U.S.A.* 105, 6741–6746.

Harle, K. M., and Sanfey, A. G. (2007). Incidental sadness biases social economic decisions in the Ultimatum Game. *Emotion* 7, 876–881.

Henson, R. N., Rugg, M. D., Shallice, T., Josephs, O., and Dolan, R. J. (1999). Recollection and familiarity in recognition memory: an event-related functional magnetic resonance imaging study. *J. Neurosci.* 19, 3962–3972.

Huettel, S. A., Mack, P. B., and McCarthy, G. (2002). Perceiving patterns in random series: dynamic processing of sequence in prefrontal cortex. *Nat. Neurosci.* 5, 485–490.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (trsnformation or not) and towards logit mixed models. *J. Mem. Lang.* 59, 434–446.

Kiehl, K. A., Laurens, K. R., Duty, T. L., Forster, B. B., and Liddle, P. F. (2001). Neural sources involved in auditory target detection and novelty processing: an event-related fMRI study. *Psychophysiology* 38, 133–142.

King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., and Montague, P. R. (2005). Getting to know you: reputation and trust in a two-person economic exchange. *Science* 308, 78–83.

Kliegl, R., Risse, S., and Laubrock, J. (2007). Preview benefit and parafoveal-on-foveal effects from word *n* + 2. *J. Exp. Psychol. Hum. Percept. Perform.* 33, 1250–1255.

Knight, R. T. (1996). Contribution of human hippocampal region to novelty detection. *Nature* 383, 256–259.

Knight, R. T., Scabini, D., Woods, D. L., and Clayworth, C. C. (1989). Contributions of temporal-parietal junction to the human auditory P3. *Brain Res.* 502, 109–116.

Koenigs, M., and Tranel, D. (2007). Irrational economic decision-making after ventromedial prefrontal damage: evidence from the Ultimatum Game. *J. Neurosci.* 27, 951–956.

Koyama, T., McHaffie, J. G., Laurienti, P. J., and Coghill, R. C. (2005). The subjective experience of pain: where expectations become reality. *Proc. Natl. Acad. Sci. U.S.A.* 102, 12950–12955.

Linden, D. E., Prvulovic, D., Formisano, E., Vollinger, M., Zanella, F. E., Goebel, R., and Dierks, T. (1999). The functional neuroanatomy of target detection: an fMRI study of visual and auditory oddball tasks. *Cereb. Cortex* 9, 815–823.

Lundstrom, B. N., Ingvar, M., and Petersson, K. M. (2005). The role of precuneus and left inferior frontal cortex during source memory episodic retrieval. *Neuroimage* 27, 824–834.

Mai, J. K., Paxinos, G., and Voss, T. (2007). Atlas of the Human Brain, 3rd Edn. New York, Academic Press.

Margulies, D. S., Kelly, A. M., Uddin, L. Q., Biswal, B. B., Castellanos, F. X., and Milham, M. P. (2007). Mapping the functional connectivity of anterior cingulate cortex. *Neuroimage* 37, 579–588.

Mealy, L., Daood, C., and Krage, M. (1996). Enhanced memory for faces of cheaters. *Ethol. Sociobiol.* 17, 119–128.

Mehl, B., and Buchner, A. (2008). No enhanced memory for faces of cheaters. *Evol. Hum. Behav.* 29, 35–41.

Mitchell, J. P. (2008). Activity in right temporo-parietal junction is not selective for theory-of-mind. *Cereb. Cortex* 18, 262–271.

Mumford, J. A., and Nichols, T. E. (2008). Power calculation for group fMRI studies accounting for arbitrary design and temporal autocorrelation. *Neuroimage* 39, 261–268.

Newbould, R. D., Skare, S. T., Jochimsen, T. H., Alley, M. T., Moseley, M. E., Albers, G. W., and Bammer, R. (2007). Perfusion mapping with multiecho multishot parallel imaging EPI. *Magn. Reson. Med.* 58, 70–81.

Oda, R. (1997). Biased face recognition in the prisoner's dilemma game. *Evol. Hum. Behav.* 18, 309–315.

Ostrowsky, K., Magnin, M., Ryvlin, P., Isnard, J., Guenot, M., and Mauguiere, F. (2002). Representation of pain and somatic sensation in the human insula: a study of responses to direct electrical cortical stimulation. *Cereb. Cortex* 12, 376–385.

Pillutla, M. M., and Murnighan, J. K. (1996). Unfairness, anger, and spite: emotional rejection of ultimatum offers. *Organ. Behav. Hum. Decis. Process.* 68, 208.

R_Development_Core_Team (2008). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, ISBN 3-900051-07-0, Available at: http://www.R-project.org.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *Am. Econ. Rev.* 83, 1281–1302.

Ranganath, C., and Rainer, G. (2003). Neural mechanisms for detecting and remembering novel events. *Nat. Rev. Neurosci.* 4, 193–202.

Rilling, J., Gutman, D., Zeh, T., Pagnoni, G., Berns, G., and Kilts, C. (2002). A neural basis for social cooperation. *Neuron* 35, 395–405.

Sanfey, A. G. (2007). Social decision-making: insights from game theory and neuroscience. *Science* 318, 598–602.

Sanfey, A. G. (2009). Expectations and social decision-making: biasing effects of prior knowledge on ultimatum responses. *Mind Soc.* 8, 93–107.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., and Cohen, J. D. (2003). The neural basis of economic decision-making in the Ultimatum Game. *Science* 300, 1755–1758.

Saxe, R., and Kanwisher, N. (2003). People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind". *Neuroimage* 19, 1835–1842.

Saxe, R., Xiao, D. K., Kovacs, G., Perrett, D. I., and Kanwisher, N. (2004). A region of right posterior superior temporal sulcus responds to observed intentional actions. *Neuropsychologia* 42, 1435–1446.

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

Shallice, T., Fletcher, P., Frith, C. D., Grasby, P., Frackowiak, R. S., and Dolan, R. J. (1994). Brain regions associated with acquisition and retrieval of verbal episodic memory. *Nature* 368, 633–635.

Singer, T., Kiebel, S. J., Winston, J. S., Dolan, R. J., and Frith, C. D. (2004). Brain responses to the acquired moral status of faces. *Neuron* 41, 653–662.

Stocker, T., Kellermann, T., Schneider, F., Habel, U., Amunts, K., Pieperhoff, P., Zilles, K., and Shah, N. J. (2006). Dependence of amygdala activation

on echo time: results from olfactory fMRI experiments. *Neuroimage* 30, 151–159.

Sutton, S., Braren, M., Zubin, J., and John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science* 150, 1187–1188.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *Q. Rev. Biol.* 46, 35–57.

Tulving, E., Markowitsch, H. J., Craik, F. E., Habib, R., and Houle, S. (1996). Novelty and familiarity activations in PET studies of memory encoding and retrieval. *Cereb. Cortex* 6, 71–79.

Tulving, E., Markowitsch, H. J., Kapur, S., Habib, R., and Houle, S. (1994). Novelty encoding networks in the human brain: positron emission tomography data. *Neuroreport* 5, 2525–2528.

van 't Wout, M., Kahn, R. S., Sanfey, A. G., and Aleman, A. (2006). Affective state and decision-making in the ultimatum game. *Exp. Brain Res.* 169, 564–568.

van 't Wout, M., and Sanfey, A. G. (2008). Friend or foe: the effect of implicit trustworthiness judgments in social decision-making. *Cognition* 108, 796–803.

Venables, W. N., and Ripley, B. D. (2002). Modern Applied Statistics with S, 4th Edn. New York, Springer.

von Neumann, J., and Morgenstern, O. (1944). Theory of Games and Economic Behavior. Princeton, NJ, Princeton University Press.

von Restorff, H. (1933). Über die wirkung von bereichsbildungen im spurenfeld. *Psychol. Res.* 18, 299–342.

Wallace, W. P. (1965). Review of the historical, empirical, and theoretical status of the von restorff phenomenon. *Psychol. Bull.* 63, 410–424.

Weiskopf, N., Hutton, C., Josephs, O., and Deichmann, R. (2006). Optimal EPI parameters for reduction of susceptibility-induced BOLD sensitivity losses: a whole-brain analysis at 3 T and 1.5 T. *Neuroimage* 33, 493–504.

Wickens, T. D. (2002). Elementary Signal Detection Theory. New York, Oxford University Press.

Xiao, E., and Houser, D. (2005). Emotion expression in human punishment behavior. *Proc. Natl. Acad. Sci. U.S.A.* 102, 7398–7401.

Yonelinas, A. P., Otten, L. J., Shaw, K. N., and Rugg, M. D. (2005). Separating the brain regions involved in recollection and familiarity in recognition memory. *J. Neurosci.* 25, 3002–3008.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.