

Distinct contributions of the amygdala and parahippocampal gyrus to suspicion in a repeated bargaining game

Meghana A. Bhatt^a, Terry Lohrenz^b, Colin F. Camerer^c, and P. Read Montague^{b,d,1}

^aBeckman Research Institute, City of Hope, Duarte, CA 91010; ^bHuman Neuroimaging Laboratory, Virginia Tech Carilion Research Institute, Roanoke, VA 24016; ^cDivision of Social Sciences and Computation and Neural Systems, California Institute of Technology, Pasadena, CA 91025; and ^dWellcome Trust Centre for Neuroimaging, University College London, London WC1N 3BG, United Kingdom

Edited by Terrence J. Sejnowski, Salk Institute for Biological Studies, La Jolla, CA, and approved April 13, 2012 (received for review January 13, 2012)

Humans assess the credibility of information gained from others on a daily basis; this ongoing assessment is especially crucial for avoiding exploitation by others. We used a repeated, two-person bargaining game and a cognitive hierarchy model to test how subjects judge the information sent asymmetrically from one player to the other. The weight that they give to this information is the result of two distinct factors: their baseline suspicion given the situation and the suspicion generated by the other person's behavior. We hypothesized that human brains maintain an ongoing estimate of the credibility of the other player and sought to uncover neural correlates of this process. In the game, sellers were forced to infer the value of an object based on signals sent from a prospective buyer. We found that amygdala activity correlated with baseline suspicion, whereas activations in bilateral parahippocampus correlated with trial-by-trial uncertainty induced by the buyer's sequence of suggestions. In addition, the less credible buyers that appeared, the more sensitive parahippocampal activation was to trial-by-trial uncertainty. Although both of these neural structures have previously been implicated in trustworthiness judgments, these results suggest that they have distinct and separable roles that correspond to their theorized roles in learning and memory.

functional MRI | game theory | neuroeconomics

Social situations often require people to assess the credibility of information communicated by others when there are grounds for suspicion about what those other people say. This requirement is especially true in competitive situations where multiple people vie for a scarce resource and must use social signals to garner information. We can roughly separate suspicion into two parts. The first part is a priori, baseline suspicion based on a person's general beliefs about people in the world and the situation at hand. The second part is the suspicion that is generated by the behavior of other people. Although this division is obviously somewhat artificial—people's baseline levels of suspicion will change over time based on their experiences in the world and people may be more or less responsive to suspicious behavior based on their baseline levels of suspicion—it is a useful starting place to start looking at how people assess the credibility of information in social situations.

The amygdala has been implicated in processing social threat in a number of situations (1, 2) and evaluating the trustworthiness of faces (1, 3–5); thus, we hypothesized that activity within the amygdala and associated structures, such as the parahippocampal gyrus, would correlate with baseline suspicion and the uncertainty generated by other people's behavior. To test this hypothesis, we investigated strategic suspicion judgments between two players, a buyer and a seller, who played 60 rounds of a bargaining task game during functional imaging.

In this bargaining game, one party, the buyer, has relevant, accurate information about the value of an object. The other party, the seller, receives a price suggestion from the buyer and

has to assess the credibility of the information to set a price for the object. The subjects interact repeatedly, allowing the seller to observe the buyer's behavior over time, but importantly, the seller does not get direct immediate feedback about the accuracy of the information that he has received. In this paradigm, sellers receive many possibly suspect suggestions from the buyer and must use these suggestions to form beliefs about both the buyer's value in any given trial and the buyer's level of credibility in general.

The buyer and seller play 60 rounds of a bargaining task (Fig. 1). At the beginning of each round, the buyer is informed of her private value v of a hypothetical object. She is then asked to suggest a price to the seller (values and prices are integers from 1 to 10). The seller then receives this suggestion and is asked to set a price p . If the seller's price is less than the private value v (which is known only to the buyer), the trade executes, and the seller receives p ; the buyer receives $v - p$, the difference between the private value and the selling price. If the seller's price exceeds the buyer's value, the trade does not execute, and both parties receive nothing. No feedback about whether the trade occurred is provided to either player.

The object has no value to either player if a trade does not occur. However, if a trade does occur, each player should prefer a sales price that favors her. Because buyers prefer lower prices and sellers prefer higher prices, this misalignment of incentives implies that the only equilibrium solution of the one-round version of this game is for no information transfer to occur (6). The buyer should babble and send suggestions with no informative relationship to her private value, and the seller should ignore this suggestion and set a price of five or six. However, this solution is the mutually optimal solution only if both players believe that the other is also playing in equilibrium (i.e., babbling is only optimal if the seller is, in fact, ignoring buyer suggestions and ignoring buyer suggestions is only optimal if they contain no meaningful information). In actuality, these types of games are often out of equilibrium, requiring people to use more involved models of belief formation such as the cognitive hierarchy style of model. In these models, people, instead, assume that they are always playing people who are less strategically sophisticated than themselves. Higher-level players form increasingly sophisticated models of how others might play, leading to a hierarchy of strategic types (6–10).

Author contributions: M.A.B., C.F.C., and P.R.M. designed research; M.A.B. and T.L. performed research; P.R.M. contributed new reagents/analytic tools; M.A.B. and T.L. analyzed data; and M.A.B., T.L., C.F.C., and P.R.M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

¹To whom correspondence should be addressed. E-mail: read@vtc.vt.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1200738109/-DCSupplemental.

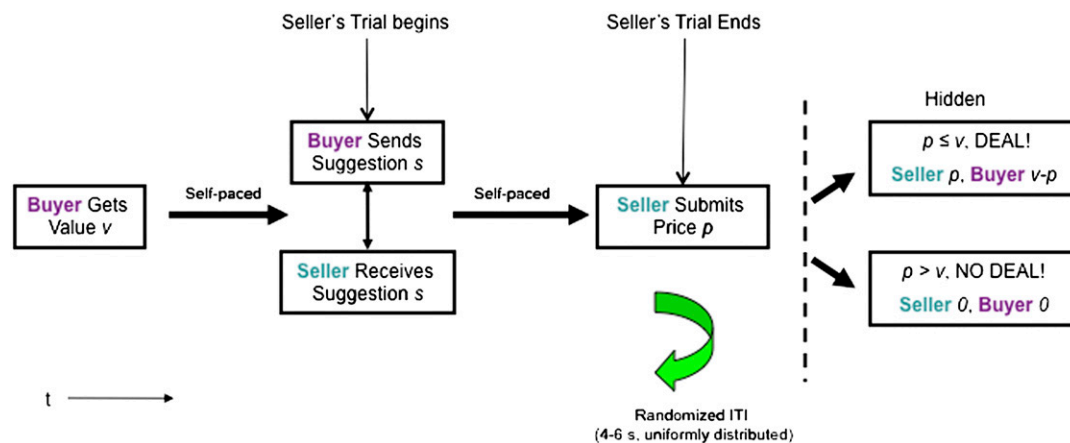


Fig. 1. Experimental task. At the beginning of each round, the computer assigns a value for the widget to the buyer. The buyer uses this information to set a final price for the object. The computer automates whether the deal occurs—if the price is less than or equal to the buyer's value v , the seller receives the price p , and the buyer receives the difference between the price and his private value, $v - p$. Otherwise, the deal fails, and neither party receives anything. Neither party is informed of the outcome of the previous trial, and payoffs are just added to a running tally of points kept by the computer.

In fact, buyers empirically vary widely in their level of credibility, with many buyers sending very reliable signals of their private values. Given the lack of feedback in the task, the sellers are forced to infer the honesty of the buyers from buyer suggestions alone. The work by Bhatt et al. (11) examined the behavior and neural activations of the buyers in this experiment and found that buyers fell into three basic behavioral groups. Incrementalists—who were relatively honest with their suggestions—and conservatives—who used the rational strategy of not revealing information—formed the bulk of the buyers (79% of the buyers) and were in roughly equal proportions. The third group, the strategists, comprised about 21% of the buyers and was actively deceptive, mimicking incrementalist behavior by sending high suggestions during low-value trials and reaping larger benefits by sending low suggestions during high-value trials. These three types corresponded roughly to levels 0, 1, and 2 players in a cognitive hierarchy model of the game. Sellers responding to these buyers were faced with the task of differentiating with whom they might be playing. Conservative behavior is relatively easy to distinguish using the stream of buyer suggestions, because suggestions from a conservative buyer generally had low variance. However, by the strategist's design, the suggestions of strategists and incrementalists are indistinguishable.

Given the relatively low percentage of strategists in the sample and noting the general human tendency to assume that opponents are likely to be less strategic than themselves, we assumed that sellers were largely concerned with distinguishing conservatives from incrementalists. In fact, in a cognitive hierarchy style model of seller behavior, the differences in predicted behavior between level 2 thinkers (who essentially assume that there are only incrementalists and conservatives) and level 3 thinkers (who acknowledge the existence of strategists) are small (*SI Materials and Methods* has details on model predictions and estimation, and *Table S1* shows CH classifications for all subjects).

Based on our assumptions, a simple proxy for sellers' assessment of buyer credibility is the SD of the suggestions received. For example, if a seller only sees one or two different suggestions over the course of the experiment, they can safely assume that the buyer suggestions contain no meaningful information and ignore them. If, however, the seller sees a wide variety of different suggestions, it is possible that those suggestions are useful. However, two sellers seeing the same stream of suggestions may still come to different conclusions about their credibility (Fig. 2*A* and *B*).

Results

Behavioral Results. We performed two separate behavioral analyses of the data: one agnostic subject-level evaluation of the behavior based on a simple regression and one model-based within-subject analysis that captured evolving beliefs about buyer credibility over time. In the first analysis, we regressed each seller's chosen prices on the buyer's suggestions. This analysis yielded three parameters of interest: the slope, intercept, and R^2 of the regression. This last parameter serves as a proxy for overall seller credulousness, with high fits indicating that sellers reliably used buyer suggestions and low fits indicating that they were not used at all. We used the SD of buyer suggestions (ζ) as a between-subject proxy for buyer-generated suspicion. R^2 and ζ were correlated ($r = 0.38$, $P < 0.001$), but this correlation was driven almost entirely by those subjects where ζ was extremely low ($\zeta < 1$), forcing a low fit. Restricting our attention to subjects who saw a greater variety of suggestions ($\zeta > 1$, $n = 64$), the correlation drops significantly ($r = 0.21$, $P = 0.09$). This relatively low correlation suggests that differences in buyer credibility alone did not adequately explain seller suspicion and that there were significant endogenous drivers of seller suspicion. To focus on these endogenous drivers of suspicion, we regressed this R^2 on ζ and let our measure of baseline suspicion (ψ) be the residuals from this regression multiplied by -1 . This measure proves to be relatively stable throughout the task. (*SI Materials and Methods*, Fig. S1).

In the second model-based analysis, we computed a parameter describing buyer-generated uncertainty about the buyer's type (i.e., the uncertainty induced by buyer's suggestions about the buyer's credibility). In this model, we assume that sellers believe that buyers are relatively naïve and send suggestions according to $s = \max(1, \min(10, [\alpha v + \epsilon]))$, where $[x]$ is the nearest integer to x . Essentially, sellers believe that buyers are sending a linearly scaled version of their true value. Notice that, in this model, the slope of the suggestion function, α , is a proxy for the credibility of the buyer. The closer that α is to zero, the less information that the seller can glean from the suggestions. Buyers with low α correspond to the conservatives described in the work by Bhatt et al. (11), whereas those buyers with higher α correspond to the incrementalists. We assume that each seller is constantly generating and updating beliefs about the credibility of the buyer based on both the stream of suggestions and the assumption that the underlying values are uniformly distributed (*SI Materials and Methods* has full details). Using this model, strategic uncertainty about buyer credibility is represented by the distribution of

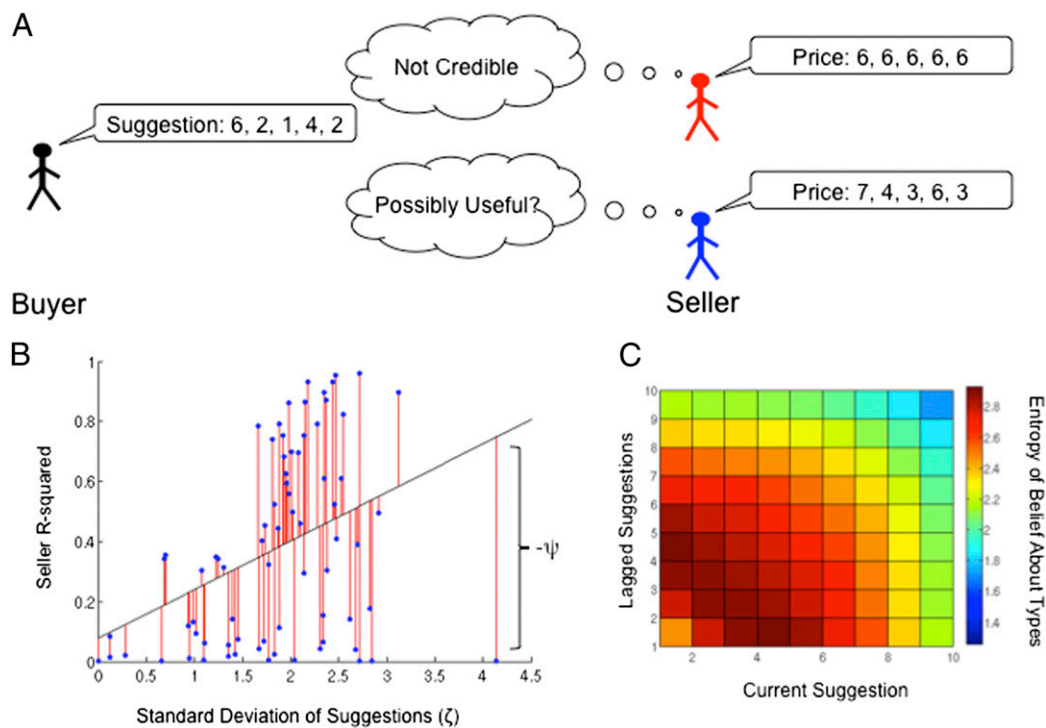


Fig. 2. (A) Although there is no feedback in this task, sellers make inferences about buyer credibility based on the stream of suggestions that they see. Two sellers seeing the same stream of suggestions may come to very different conclusions based on their a priori beliefs about how trustworthy buyer suggestions are likely to be. A suspicious seller (red) will generally ignore the buyer's suggestion, whereas an unsuspecting seller seeing the same suggestions (blue) will tend to base their chosen prices on the buyer's suggestions. (B) Empirically, sellers seeing similar streams of suggestions, as measured by the SD of those suggestions (ζ), showed widely varying behavior, as measured by the R^2 of the regression of the seller's chosen prices on the buyer's suggestions. The scatter plot shows that many sellers showed near zero R^2 values despite seeing highly variable suggestions, whereas others displayed fits approaching one. The red lines represent the residuals of the R^2 regressed on ζ , and we multiplied this quantity by -1 to get ψ , our measure of baseline suspicion. (C) We modeled how sellers should rationally make inferences about buyer credibility based on the buyer's current and latest suggestion. We used the entropy of their beliefs about the buyer's type in any given trial as a measure of buyer-generated uncertainty. Uncertainty is minimized when the buyer sends high suggestions, implying their relative credibility. Interestingly, uncertainty is maximized when buyers send one low and one intermediate suggestion, because two low suggestions can actually make the seller relatively certain that the buyer is untrustworthy.

seller's beliefs over α (ranging from credible at $\alpha = 1$ to babbling at $\alpha = 0$). We used the entropy of this distribution as a measure of the seller's uncertainty about the buyer's type in each trial. We calculated these entropies assuming limited memory based only on the current and previous trials' suggestions. Fig. 2C shows a heat map representation of this measure based on every possible combination of previous and current trial suggestions. Notice that strategic uncertainty about buyer type is minimized when sellers see a high suggestion, implying that they are likely to be relatively credible, but it is also relatively low when they only see low suggestions, implying that they are definitely untrustworthy.

In addition to these quantitative analyses, we collected debriefing from all subjects—asking them how they chose to approach the task. Many sellers reported being highly suspicious of buyer suggestions (multiple sellers used the phrase *lowball* when discussing buyers' suggestions). For example, one seller wrote, "I assumed I could NOT trust the buyer and that he might try to rip me off at any chance he had." Unsurprisingly, this seller showed relatively high ψ despite playing a fairly credible buyer (0.26; among the top 25% of sellers in baseline suspicion). Another suspicious seller reported assuming that the buyer was always "lying." Conversely, one of the more trusting players stated, "I took a low-risk approach on the idea that the buyer would not bid at or above the price . . ." Unlike the suspicious sellers, he did not use words like "trust" or "lying" anywhere in his description. Importantly, no subjects expressed suspicion that they might not be playing another human being, with most referring to the buyer as "he" or "she."

Functional MRI Results. The amygdala has been associated with fear, especially fear generated by socially threatening cues, and the evaluation of trustworthiness (5). We hypothesized that amygdala activity should be modulated by both buyer-generated and baseline suspicions. To test the hypothesis, we regressed neural activity over the decision epoch on both ψ and ζ for each subject and applied small-volume corrections in the 20-mm spheres around the foci of amygdala activation found in the study by Winston et al. (5) of the evaluation of trustworthy faces. Bilateral dorsal amygdala activation correlated strongly with ψ [$P < 0.001$ uncorrected, $P < 0.05$ corrected for the familywise error in the 20-mm spheres around (18, 0, -24) and (-16, -4, -20), respectively]. These results indicate that amygdala activity correlates more with internally generated, a priori levels of suspicion than externally cued suspicion (Fig. 3). No clusters correlated to ζ survived correction for multiple comparisons either at the whole-brain level or with the small-volume correction (Tables S2 and S3 are complete activation tables).

Within-subject analysis showed that bilateral parahippocampus activity correlates with our trial-level measures of buyer-generated uncertainty at trial onset. Specifically, the less certain that a seller was about a buyer's type, the more active that the parahippocampus was; this bilateral activation survived whole-brain correction for multiple comparisons at the 0.05 level (cluster correction using an uncorrected threshold of $P < 0.001$). Region of interest analyses on these two clusters reveal that the degree to which these areas correlate with buyer-generated uncertainty, as measured by the average regression coefficient on trial-by-trial entropy over the cluster, is significantly negatively

The current study shows a significant between-subject difference in amygdala activation based on how suspicious subjects were of buyer suggestions. We find that amygdala activation correlates with baseline suspicion, which is consistent with the lesion literature showing impaired fear responses in people with amygdala damage (1, 2, 14, 21) (i.e., people with lower average amygdala activity were less suspicious overall), even controlling for the implied credibility of the suggestions that they saw. Extrapolating, one might predict that subjects with amygdala damage would show no suspicion whatsoever. Notice that, if we were to, instead, consider amygdala activation as a signal of nonsocial risk aversion, we might expect the opposite result; if sellers believed, as they seemed to do, that most buyers were suggesting prices under the true value, using those suggestion would ensure a transaction with a positive, low-risk payoff in every round. In fact, as noted above, one of the least suspicious sellers took this approach, calling it the low-risk strategy.

Activation in the parahippocampal gyrus and other areas of the medial temporal lobe has generally been implicated in memory formation and learning (22). The area has dense interconnectivity with the amygdala, which has been theorized to interact with the hippocampal complex, including the parahippocampus, in the formation of emotional memories (23, 24). In fact, the amygdala and hippocampal complex have been theorized to subserve the formation of implicit/nondeclarative and explicit/declarative memory, respectively (25–27).

In this study, we have shown that the parahippocampal gyrus and amygdala correlate with distinct aspects of trustworthiness judgments. Although the amygdala correlates strongly with baseline suspicion (arguably the more emotional aspect of trustworthiness judgments), the parahippocampal gyrus correlates more strongly with the rational trial-by-trial uncertainty generated by another individual's behavior. Additionally, this response to trial-by-trial uncertainty is enhanced in sellers facing buyers that seem more untrustworthy overall. The work by Sharot et al. (28) found a similar dissociation between these two areas when studying the feeling of remembering. In their study, amygdala activity was enhanced for memories of emotional images, whereas activity in the parahippocampal gyrus was enhanced for memories of neutral images. Similarly, in our study, amygdala activity is enhanced for baseline suspicion rather than the uncertainty generated by another person's behavior, whereas activity in the parahippocampal gyrus is enhanced during rational updating of beliefs about buyer behavior, requiring retrieval of the buyer's recent signals.

In this work, we have extended the literature about judgments on trustworthiness to distinguish between simple baseline suspicion and the ongoing learning process around uncertainty generated by another person's behavior. The more emotional factor of baseline suspicion is associated with activity in the amygdala, consistent with the literature tying this area to fear processing.

However, the more rational factor of buyer-generated uncertainty is associated with activity in the parahippocampal gyrus, consistent with its role in memory retrieval and learning.

Materials and Methods

Subject Information. Subjects were recruited from a large database maintained by the Human Neuroimaging Laboratory. Most subjects were affiliated with the Medical Center or Rice in various ways. Subjects were recruited in pairs and kept separate throughout the task, and therefore, the experiment was completely anonymous. Subjects were informed that they were playing a real person and seemed to believe this case to be true.

After filling out a standard screening form and giving consent, they were given written instructions on the task. After reading the instructions, they were walked through the task again verbally and asked to answer a few questions about hypothetical bargaining rounds to make sure that they fully understood the task. If any of these questions was answered incorrectly, the mistake was explained, and another scenario was presented.

After both subjects were fully instructed, they were loaded into the scanner and participated in 60 rounds of the bargaining task. At the end, they were told their total earnings and given an open-ended debriefing sheet on which they were asked to describe, in their own words, how they had approached the task. Seventy-six pairs of subjects were scanned. However, because of a mechanical error, functional MRI (fMRI) data were only acquired for 74 of the sellers. Data are reported for this subset. Mean age for these sellers was 29 y, SD was 7.6 y, and 45 of 74 subjects were female.

fMRI Methods. fMRI data were collected using 3-T Siemens scanners on 74 healthy subjects recruited in accordance with a protocol approved by the Baylor College of Medicine Institutional Review Board. High-resolution T1-weighted scans were acquired using an MP-Rage sequence. Functional images were acquired with repetition time (TR) = 2,000 ms and echo time (TE) = 25 ms; 37 4-mm slices were acquired 30° off the anteroposterior commissural line, yielding functional voxels that were 3.4 × 3.4 × 4 mm.

Data were preprocessed using SPM2 algorithms for slice-timing correction, motion correction, coregistration, gray/white matter segmentation, and normalization to the Montreal Neurological Institute template. Functional images were smoothed spatially using an 8-mm Gaussian kernel. All data were high pass-filtered (128 s); the regression error structure was assumed to be autoregressive with order 1 [AR(1)]. Postprocessing voxels were 4 × 4 × 4 mm.

We considered two general linear models on the data. Key presses, head motion, and time derivatives were included as nuisance regressors in both models. The first model used a boxcar regressor beginning at trial onset and ending at decision parameterized by both the suggestion received and the chosen price in each trial. The second model used separate point regressors at trial onset and decision that were both parameterized by the entropy of seller beliefs about buyer type. Regressors were convolved with the standard hemodynamic response function. After regions of interest were identified from the whole-brain analysis, time series were extracted in each cluster and averaged to produce time courses anchored to events of interest.

ACKNOWLEDGMENTS. This research was funded by National Institutes of Health Grants 1 RC4 AG039067, R01 DA11723-02, and 2R01 MH085496-05A2, by The Kane Foundation Fellowship, and by the Wellcome Trust. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

- Adolphs R, Tranel D, Damasio AR (1998) The human amygdala in social judgment. *Nature* 393:470–474.
- Adolphs R (2008) Fear, faces, and the human amygdala. *Curr Opin Neurobiol* 18:166–172.
- Todorov A, Engell AD (2008) The role of the amygdala in implicit evaluation of emotionally neutral faces. *Soc Cogn Affect Neurosci* 3:303–312.
- Engell AD, Haxby JV, Todorov A (2007) Implicit trustworthiness decisions: Automatic coding of face properties in the human amygdala. *J Cogn Neurosci* 19:1508–1519.
- Winston JS, Strange BA, O'Doherty J, Dolan RJ (2002) Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nat Neurosci* 5:277–283.
- Crawford VP, Sobel J (1982) Strategic information transmission. *Econometrica* 50:1431–1451.
- Costa-Gomes M, Crawford VP, Broseta B (2001) Cognition and behavior in normal-form games: An experimental study. *Econometrica* 69:1193–1235.
- Nagel R (1995) Unraveling in guessing games: An experimental study. *Am Econ Rev* 85:1313–1326.
- Camerer CF, Ho TH, Chong J (2004) A cognitive hierarchy model of games. *Q J Econ* 119:861–898.
- Coricelli G, Nagel R (2009) Neural correlates of depth of strategic reasoning in medial prefrontal cortex. *Proc Natl Acad Sci USA* 106:9163–9168.
- Bhatt MA, Lohrenz T, Camerer CF, Montague PR (2010) Neural signatures of strategic types in a two-person bargaining game. *Proc Natl Acad Sci USA* 107:19720–19725.
- Phelps EA (2006) Emotion and cognition: Insights from studies of the human amygdala. *Annu Rev Psychol* 57:27–53.
- Hsu M, Bhatt M, Adolphs R, Tranel D, Camerer CF (2005) Neural systems responding to degrees of uncertainty in human decision-making. *Science* 310:1680–1683.
- De Martino B, Camerer CF, Adolphs R (2010) Amygdala damage eliminates monetary loss aversion. *Proc Natl Acad Sci USA* 107:3788–3792.
- Bechara A, Damasio H, Damasio AR, Lee GP (1999) Different contributions of the human amygdala and ventromedial prefrontal cortex to decision-making. *J Neurosci* 19:5473–5481.
- De Martino B, Kumaran D, Seymour B, Dolan RJ (2006) Frames, biases, and rational decision-making in the human brain. *Science* 313:684–687.
- Coricelli G, et al. (2005) Regret and its avoidance: A neuroimaging study of choice behavior. *Nat Neurosci* 8:1255–1262.
- Morris JS, et al. (1996) A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature* 383:812–815.

19. de Gelder B (2006) Towards the neurobiology of emotional body language. *Nat Rev Neurosci* 7:242–249.
20. Heberlein AS, Adolphs R (2004) Impaired spontaneous anthropomorphizing despite intact perception and social knowledge. *Proc Natl Acad Sci USA* 101:7487–7491.
21. Glascher J, Adolphs R (2003) Processing of the arousal of subliminal and supraliminal emotional stimuli by the human amygdala. *J Neurosci* 23:10274–10282.
22. Paller KA, Wagner AD (2002) Observing the transformation of experience into memory. *Trends Cogn Sci* 6:93–102.
23. Phelps EA (2004) Human emotion and memory: Interactions of the amygdala and hippocampal complex. *Curr Opin Neurobiol* 14:198–202.
24. Kilpatrick L, Cahill L (2003) Amygdala modulation of parahippocampal and frontal regions during emotionally influenced memory storage. *Neuroimage* 20:2091–2099.
25. Bechara A, et al. (1995) Double dissociation of conditioning and declarative knowledge relative to the amygdala and hippocampus in humans. *Science* 269:1115–1118.
26. Milner B, Squire LR, Kandel ER (1998) Cognitive neuroscience and the study of memory. *Neuron* 20:445–468.
27. Knight DC, Waters NS, Bandettini PA (2009) Neural substrates of explicit and implicit fear memory. *Neuroimage* 45:208–214.
28. Sharot T, Delgado MR, Phelps EA (2004) How emotion enhances the feeling of remembering. *Nat Neurosci* 7:1376–1380.