Neural correlates of mentalizing-related computations during strategic interactions in humans

Alan N. Hampton†, Peter Bossaerts†‡, and John P. O’Doherty†§

†Computation and Neural Systems Program and ‡Division of Humanities and Social Sciences, California Institute of Technology, 1200 East California Boulevard, MC 228-77, Pasadena, CA 91125

Edited by Edward E. Smith, Columbia University, New York, NY, and approved February 20, 2008 (received for review November 22, 2007)

Competing successfully against an intelligent adversary requires the ability to mentalize an opponent’s state of mind to anticipate his/her future behavior. Although much is known about what brain regions are activated during mentalizing, the question of how this function is implemented has received little attention to date. Here we formulated a computational model describing the capacity to mentalize in games. We scanned human subjects with functional MRI while they participated in a simple two-player strategy game and correlated our model against the functional MRI data. Different model components captured activity in distinct parts of the mentalizing network. While medial prefrontal cortex tracked an individual’s expectations given the degree of model-predicted influence, posterior superior temporal sulcus was found to correspond to an influence update signal, capturing the difference between expected and actual influence exerted. These results suggest dissociable contributions of different parts of the mentalizing network to the computations underlying higher-order strategizing in humans.

H umans, like many other primates, live in a highly complex social environment in which it is often necessary to interact with, and compete against, other individuals to attain reward. Success against an intelligent adversary with competing objectives likely depends on the capacity to infer the opponent’s state of mind, to predict what action the opponent is going to select in future, and to understand how an individual’s own actions will modify and influence the behavior of one’s opponent. This ability is often referred to as “mentalizing” and has been linked to a number of specific regions thought to be specifically engaged when processing socially relevant stimuli, and especially when inferring the state of minds of others (1). Neuroimaging studies in humans have implicated a specific network of brain regions including dorsomedial prefrontal cortex (PFC), posterior superior temporal sulcus (STS), and the temporal poles (2,3) while subjects engage in tasks relevant to mentalizing, such as evaluating false beliefs or social transgressions (4, 5), describing the state of biological movements (6–8), and playing interactive games (9–11). However, although these studies have provided insight into what brain regions may be involved in the capacity to mentalize, the question of how this function is implemented at the neural level has received relatively little attention to date.

The goal of the present study was to build a simple model describing computations underlying the capacity to mentalize (in the context of a strategic game) and to determine whether different components of this model were correlated with neural activity in parts of the mentalizing network. To assess competitive interactions experimentally, we studied pairs of human subjects while they played each other in a two-player strategic game called the “inspection” game (or generalized matching pennies), in which opponents have competing goals (Fig. 1A and B). One of the players was being scanned with functional MRI (fMRI), and the opponent was playing outside the scanner. The “employer” could either “inspect” or “not inspect,” and the “employee” could either “work” or “shirk.” The employer received 100 cents if he/she did not inspect and the employee worked and 25 cents if he/she inspected and caught the employee shirking. Otherwise he/she got zero cents. In contrast, the employee got 50 cents for working when the employer inspected and for shirking when the employer did not inspect, otherwise getting zero cents as well. Both players had competing objectives, in that when one player won in a given trial, the other one lost.

A player can in principle use a number of different strategies to try to win in such a game. Perhaps the simplest strategy is on each trial to simply choose the action that in the recent past gave the most reward. This strategy is referred to as reinforcement learning (RL) and approximates the optimal solution for many different types of decision problem in nonstrategic contexts, even for decision problems with complex higher-order structure whereby such structure can be accommodated by a sufficiently nuanced model of the state-space and transition probabilities (12, 13). However, such a strategy would be devastating for an individual in a competitive scenario because a clever opponent could detect the regularity in the reinforcement learner’s choices to work out what action the reinforcement learner is going to choose next and exploit that knowledge by choosing the confounding response.

A more sophisticated approach is to try to predict the opponent’s next actions by taking into account the history of prior actions by the opponent and then choosing the best response to that predicted action, a strategy known as “fictitious play” (14–16). A fictive learner is, in contrast to a reinforcement learner, employing an elementary form of mentalizing, because they are engaging a representation of the actions and intentions of their opponent.

However, an even more cognitively sophisticated and Machiavellian strategy a player could use in this game is to not only track the opponent’s actions, but also to incorporate knowledge of how one’s own actions influence the opponent’s strategy. Simply put, this involves a player’s building a prediction of what the opponent will do in response to the player’s own actions. For example, the more the employer inspects, the higher the probability the employee will work in subsequent trials. The employer can then use this knowledge to make choices with higher expected rewards in subsequent trials, i.e., not inspect. We will term this strategy the “influence” learning model (see Table 1 for a comparison of the different models).

Although the behavioral game theory literature has demonstrated that humans think strategically in one-shot games [they consider what the opponent could possibly believe (17, 18)], the updating of beliefs about the beliefs of the opponent has rarely been incorporated in the analysis of learning in repeated play. Our modeling approach differs from the two exceptions (19, 20) in both computational modeling | decision making | functional MRI | neuroeconomics

Author contributions: A.N.H., P.B., and J.P.O. designed research; A.N.H. performed research; A.N.H. analyzed data; and A.N.H., P.B., and J.P.O. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

†To whom correspondence should be addressed. E-mail: jdoherty@hss.caltech.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0711099105/DCSupplemental.

© 2008 by The National Academy of Sciences of the USA

www.pnas.org/cgi/doi/10.1073/pnas.0711099105

PNAS May 6, 2008 vol. 105 no. 18 6741–6746
The RL model updates the value of the chosen action a with a simple Rescorla–Wagner (35) prediction error (\(R_t - V_t\)) as the difference between received rewards and expected rewards, where \(\eta\) is the learning rate. The fictitious play model instead updates the state (strategy) of the opponent \(p_t^i\) with a prediction error (\(p_t - p_t^i\)) between the opponent’s action and expected strategy. The influence model extends this approach by also including the influence (\(Q_t - q_t^i\)) that a player’s own action \(Q_t\) has on the opponent’s strategy (see Methods).

its simplicity (fewer parameters are needed) and its form, which is more relevant for neuroscience, because it disentangles the two components of the prediction error, as we explain later on.

We chose the inspection game for our experimental study because there is no simple best way of playing it (in game-theoretic terms, it does not have a pure-strategy equilibrium): always inspecting makes the employee always work, in which case inspecting is not optimal, etc. Hence, regardless of the learning model implemented, strategies will keep switching, and, consequently, learning never disappears.

Results

Model Fits to Behavior. To address which of the above strategies most closely captured subjects’ behavior, we fit each model to behavior separately and compared the goodness of fit of each model. We found that the influence model provided a significantly better fit to subjects’ behavior (see Methods). We found that the influence model provided a significantly better fit to subjects’ behavior (\(\eta\)) than did either the fictitious play rule or the RL rule, even when taking into account the different number of free parameters in each model by performing an out-of-sample test (Fig. 1C and supporting information (SI) Fig. S1). The influence model also fit better than a variation of the experience-weighted attraction (EWA) learning rule, which involves a combination of RL and fictitious play but has the same number of parameters and hence is equal in model complexity to the influence model (15). Fig. 1D shows the relationship between the probability of an action being selected as predicted by the influence model, and actual subject choices. These findings suggest that subjects are not only using representations of the opponents’ future choices to guide choice, but are also employing representations of the opponents’ likely responses to their own actions.

fMRI Correlates of Influence-Based Expectations. We next analyzed the fMRI data from the player being scanned to determine whether we could find evidence of neural signals reflecting the different components of the influence model, and, if so, whether those signals are better accounted for by the influence model than by the fictitious play or simple RL models. A comparison of brain signals associated with the expected reward of the chosen action, as predicted by each model, is shown in Fig. 2A. Expected value signals from the influence model were significantly correlated with neural activity in medial orbitofrontal cortex (mOFC), medial PFC (mPFC) (encompassing both ventral and dorsal aspects, significant at \(P < 0.05\) corrected for small volume (SVC), and right temporal pole (\(P < 0.05\) SVC). By contrast, only weak correlations with the expected value signals from the fictitious play model were found in mOFC, whereas no significant correlations were found with expected value as computed by the simple RL model.

Comparison of Influence, Fictitious, and RL Model Fits to fMRI Data. We tested for brain regions showing a significantly better fit to the influence model than the RL model. This analysis revealed significant effects extending from mid to dorsal mPFC (\(P < 0.05\) SVC; Fig. 2B), as well as in the right temporal pole (Fig. S2). The regression fits of the three models are shown in Fig. 2C for mPFC, demonstrating the superiority of the influence model in accounting for neural activity in this area. We then binned BOLD activity from mPFC according to the expected reward as predicted by the influence model to illustrate the relationship between evoked fMRI responses and the model predictions (Fig. 2D). These data show that the influence model provides a significantly better account of the neural data in mPFC than does a simple RL model. In addition to the voxel-based analysis we performed an ROI analysis by

Table 1. Model update rules

<table>
<thead>
<tr>
<th>Model</th>
<th>Update rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>RL</td>
<td>(V_{t+1}^i = V_t^i + \eta(R_t - V_t^i))</td>
</tr>
<tr>
<td>Fictitious</td>
<td>(p_{t+1}^i = p_t^i + \eta(p_t - p_t^i))</td>
</tr>
<tr>
<td>Influence</td>
<td>(p_{t+1}^i = p_t^i + \eta(p_t^i - p_t^i) - \kappa(Q_t - q_t^i))</td>
</tr>
</tbody>
</table>

Fig. 1. Inspection game and behavioral results. (A) Two interacting players are each individually given two action choices at the beginning of each trial. (B) Game payoff matrix for the inspection game used in this paper. (C) Log likelihood errors for each computational model tested shows that the influence model, which incorporates the effects of players’ actions influencing their opponents, has a better fit to subjects’ behavior than either the RL or fictitious play models or these two models combined. To account for overfitting and the effects of differences in free parameters between models we used an out-of-sample prediction validation technique, as shown in Fig. S1. Error bars show the SEM of individual log likelihoods. (D) Furthermore, the actual probability of a player taking a specific behavioral action is linear with respect to the probability of choosing that action as computed by the influence model. Here, behavior and predictions are shown separately for the employer and employee. Error bars are SEM over subjects.
extracting averaged activity across voxels within an 8-mm sphere in mPFC [centered on mean coordinates from a metaanalysis by Frith and Frith (3)]. We fit the expected reward regressor from the RL and the influence model separately to these data and compared the regression fits across subjects using a paired t test. The expected reward signal from the influence model provided a significantly better fit than the expected reward signal from the RL model at P < 0.005 within our mPFC ROI, confirming the conclusions from our voxel-based analysis.

We then aimed to differentiate between the effects of the influence model and the more closely related fictitious play model in this area. For this, we looked specifically at the points in the experiment when the predictions of these two models differ. In particular, the influence model predicts that the expected value after a switch in action choice (i.e., moving from working to shirking or vice versa on successive trials) is on average higher than the expected reward when not switching choice (i.e., taking the same action on successive trials), whereas the fictitious play and indeed RL models predict exactly the opposite (Fig. 2E). This effect is greatest for the employee, because behavioral fits indicate that subjects exert more influence on their opponent when playing this role. An analysis of BOLD activity in the mPFC region of interest at the time of choice revealed a positive signal in this area on switch compared with nonswitch trials for the employee, consistent with the predictions of the influence model but not with either the fictitious play or simple RL models (Fig. 2F). These results therefore suggest that the influence model does indeed account better for neural activity in mPFC than the fictitious play model.

Fig. 2. Expected reward signals. (A) At the time of choice, the expected reward of the action selected by a player is shown across the brain as calculated by different computational models. The expected reward signal from the influence model is correlated significantly with BOLD responses in mOFC (0, 36, −21 mm, z = 3.56), mPFC (−3, 63, 15 mm, z = 3.29), and in the right temporal pole (42, 15, −39, z = 3.98), the latter two areas surviving at P < 0.05 correction for small volume (SVC) within an 8-mm sphere centered on coordinates from areas implicated in mentalizing (3), whereas only the fictitious play model has significant activity in mOFC (at P < 0.001). The RL model had no significant activity correlating with expected reward anywhere in the brain. (B) An analysis to test for areas showing neural activity related to expected reward, which is explained significantly better by the influence model than by the RL model, revealed statistically significant effects in mPFC (−3, 57, 12 mm, z = 3.11; P < 0.05 SVC). (C) The average correlation coefficients for each model from the area reported in B (extracted from all voxels showing effects at P < 0.005 in mPFC). All images shown depict whole-brain voxel-wise comparisons; small volumes are defined only for the purposes of correction for multiple comparisons. (D) fMRI activity in mPFC shows a linear relation with binned expected reward probabilities as computed by the influence model (fMRI activity extracted from individual peaks in a 10-mm search radius centered on peak from B). (E) The computational models tested in this article make distinctly different predictions about the expected reward signals after switching actions (switch) or sticking to the same action (nonswitch) as a consequence of influencing the opponent. Intuitively, the underlying reason is that both RL and fictitious play will most likely “stay” after a reward and “switch” after a nonreward. However, the influence model has a higher incentive to switch even after receiving a reward. That is, expected reward signals associated with a specific action do not necessarily increase after the receipt of a reward when taking into consideration the influence that specific action exerts on the opponent’s strategy. (F) fMRI responses in mPFC at the time of choice on switch compared with nonswitch trials show a response profile consistent with the influence model and not the fictitious play models or RL models (the data are extracted from a 10-mm sphere centered on peak from B). The difference between the employee and employer was significant at P = 0.02.
Influence signals in the brain. (A) At the time of outcome, the influence update of the inferred opponent’s strategy shows significant correlations with activity in STS bilaterally (−57, −54, 0 mm, _z_ = 3.32 and 60, −54, 9 mm, _z_ = 3.35; _P_ < 0.05 SVC). (B) The degree to which a subject thinks he/she is influencing his/her opponent can be measured by taking the difference in log-likelihood fits between the influence and fictitious models on each player’s behavior. Likewise, brain regions involved in computing the influence on the opponent will correlate more strongly with the influence model for subjects invoking this approach when compared with subjects that do not. Influence signals were found to significantly covary with the model likelihood difference (influence − fictitious) across subjects in mPFC (−3, 51, 24 mm, _z_ = 4.09; _P_ < 0.05 SVC). (Right) The relationship between influence regression coefficients and model likelihood differences in mPFC. All images shown depict whole-brain voxel-wise comparisons; small volumes are defined only for the purposes of correction for multiple comparisons.

The influence update signal ( _P_ < 0.05 SVC; Fig. 3A), suggestive of a role for this region in guiding the update of expected value representations in mPFC. Prediction error signals were found to correlate with neural activity in ventral striatum bilaterally (see Fig. S3), consistent with many previous findings implicating this area in prediction error coding (21–24). Moreover, this analysis revealed significant prediction error effects in mPFC, suggesting that this signal could also contribute to the updating of expectations in this region.

**fMRI Correlates of Between-Subject Differences in Influence-Related Strategizing.** To further investigate differences between the influence and fictitious play models, we examined between-subject variability in the degree to which the influence model provided a better fit to subjects’ behavior than the fictitious play model, by comparing the difference in the likelihoods of the two models and correlating that with neural activity elicited by the influence update signal. This measure can be taken as an assay of the individual differences in the degree of influence-based strategizing within our subject group. We found a significant between-subject correlation in the degree of influence activity and the difference in likelihoods between the influence and fictitious play models in dorsomedial PFC ( _P_ < 0.05 SVC; Fig. 3B). These results suggest that, among subjects who strategize more, the influence-based model correlated better with neural activity in mPFC.

**Correlations Between Regions During Task Performance.** Given the structure of our computational model, an important implication of our findings is that neural activity in mPFC ought to be predictable from a combination of the signals contained in posterior STS (pSTS) and ventral striatum. To test this hypothesis we computed correlations between activity in mPFC and activity in pSTS and ventral striatum separately for each different time point within a trial. We found a significant increase in correlations between activity in STS and mPFC and between activity in ventral striatum and mPFC after receipt of the outcome (when prediction errors and influence errors are generated) compared with the period before the outcome was delivered (significant at _P_ < 0.01 for both regions; Fig. 4). Moreover, activity in mPFC at the time of outcome was significantly better predicted by a linear combination of signals in STS and ventral striatum than by the signals in either of these regions alone (at _P_ < 10⁻⁶). These findings therefore support the possibility that mPFC, ventral striatum, and pSTS constitute a functionally interacting network underlying computations in strategic game playing, consistent with the tenets of our computational model.
Discussion

In the present study we show that a model that captures an updating strategy in which individuals keep track not only of the actions of the opponent, but also of how opponents are influenced in response to their own actions, provides a good account of behavior during performance of a simple strategic game. We also show that specific computational signals needed for the implementation of such a strategy are correlated with neural activity in different parts of the mentalizing network.

mPFC and pSTS were found to fulfill very different roles in the context of our computational framework. Whereas activity in mPFC was found to track the predicted future reward corresponding to a particular choice given the degree of influence expected, activity in pSTS was found to correspond to an update signal, capturing the difference between the degree of influence expected on a given trial and the actual influence exerted once the outcome had been revealed.

Reward expectations in mPFC were found to be correlated specifically with the predictions of the full influence model, suggesting that these predictions take into account the subjects’ expectations of the degree of influence that would be exerted on the opponent given the subject’s own inference of how the opponent would respond to his or her own actions. By contrast, reward expectations in this region were not captured well by either a fictitious play model, which simply tracks the actions of the opponent without considering the opponent’s reactions to the subject’s own actions, or by RL, which tracks only the reward expected given previous choices of the same action. These findings suggest that representations of expected reward in this region take into account inferences about the intentions or beliefs of the opponent toward oneself, often considered a hallmark of the psychological construct of mentalizing.

Another component of the mentalizing network, pSTS, was found to be correlated with the influence that a player’s action had on the opponent’s strategy. This area has previously been implicated in processing stimuli related to living agents and biologically relevant motion (6, 25). Here we provide evidence that this region is involved in updating an individual’s strategy based on computations related to the degree of influence an individual has exerted on their opponent during strategic social interactions. Our computational modeling approach also allowed us to separate out prediction error signals arising from simple RL from those arising from the more complex influence updating mechanism, because the update of expectations in the full influence model is accomplished by a combination of these two signals. When we tested for the presence of prediction error signals arising from the RL component of the model we found significant correlations with those signals in the ventral striatum bilaterally, consistent with many previous reports (21, 26, 27). This signal is independent and dissociable from the influence update signal present instead in pSTS. Taken together these findings suggest that two distinct updating mechanisms are present in the human brain at the same time during strategic interactions: those relating to the difference between the expected and actual rewards (the RL prediction error) and those related to the difference between expected and actual influence exerted.

We also explored functional correlations among the three key brain regions identified as containing signals relevant to our influence model to test an important implication of our model. Namely, activity in mPFC particularly at the time that the outcome is revealed within a trial ought to be correlated with a linear combination of activity in the two regions containing the update signals: ventral striatum and pSTS. Consistent with the predictions of our model, we found a significant increase in correlations between activity in both ventral striatum and pSTS with activity in mPFC at the time the outcome was revealed and the error signals were generated, compared with other time points in the trial. Moreover, activity in mPFC was better predicted by a linear combination of activity in these two regions than by activity in one or the other region alone. These findings support the conclusion that pSTS, ventral striatum, and mPFC constitute a functionally interacting network for implementing the computations relevant to mentalizing.

Although in the present study players understand the effects of influencing the opponent, a key outstanding issue is how they could use that knowledge to alter the opponent’s behavior so as to receive bigger future rewards, such as reputation building and teaching (20), or Stackelberg strategies, in which one player commits to a certain strategy and forces the other player to follow suit (28). Furthermore, although in the present study human players always faced real human opponents, an interesting question for further study would be whether similar mechanisms are engaged in these areas when subjects are playing an intelligently adaptive but non-human computer, a manipulation often used when probing “theory of mind” areas in human imaging studies (9–11). Another open question is whether other animals besides humans have the capacity for sophisticated strategic computations of this sort, or whether the capacity to engage in such high-level strategies is a uniquely human trait. Although previous studies of strategic game playing in rhesus macaques indicate that these animals do use simple RL and possibly fictitious updating (29–31), it has not yet been addressed whether they are capable of higher-level strategizing as found here in human subjects.

In this study we have taken the first steps in attempting to characterize the neural underpinnings of mentalizing during strategic interactions in terms of a simple computational model. mPFC and pSTS made distinct functional contributions. Whereas activity in mPFC was found to track the predicted future reward corresponding to a particular choice given the degree of influence expected, activity in pSTS was found to correspond to an influence update signal, capturing the difference between the degree of influence expected on a given trial and the actual influence known to have occurred once the outcome had been revealed. Accordingly, whereas signals in mPFC relating to expectations may be used to guide choice during game performance, signals in pSTS may be used to modulate or change influence expectations on the basis of the actual outcomes experienced. These areas have previously been implicated in mentalizing and in theory of mind but have never been shown to have correlations with distinct computational processes that may potentially underlie such capacities. More generally, the present results show how the application of quantitative computational models to neuroimaging and behavioral data can be used not only to advance knowledge of simple learning situations but also to unlock the complexities of social and strategic interactions (32, 33).

Methods

Subjects. Thirty-two healthy normal subjects participated in this study, of which 16 (25 ± 1 year old, seven female) were scanned while playing a competitive game in pairs with the other 16. Subject pairs were prescreened to make sure that the subjects in each pair did not know each other before the experiment to reduce the possibility of collusion. However, data from one pair of subjects was in fact discarded because of evidence of collusion during the game. The subjects were also preassessed to exclude those with a prior history of neurological or psychiatric illness. All subjects gave informed consent, and the study was approved by the Institute Review Board at California Institute of Technology.

Task. Functional imaging was conducted by using a Siemens 3.0 Tesla Trio MRI scanner to acquire gradient echo T2*-weighted echoplanar images. Each pair of subjects played three game sessions. One subject used a computer terminal and keyboard to play the game while the other was in the scanner using goggles as visual input (Resonance Technologies) and a button box to choose an action. The first session, of 50 trials, was for training; the second two sessions, of 100 trials each, are reported in this article. Player roles alternated between the two subjects in each session. Thus, the scanned subjects reported in this paper played both roles in subsequent sessions (employer and employee). We also included randomly intermixed null event trials, which accounted for 33% of the total number of trials in a session. These trials consist of the presentation of a fixation cross for 7 s. Before entering the scanner, subjects were informed that they would receive...
what they earned in a randomly selected session (not training), plus an additional $10. The payoff matrix’s association with players’ roles and actions (Fig. 1A and B) remained fixed throughout an experiment but was rotated across experiments to create a balanced design with respect to roles and role actions. However, for convenience, we refer to the roles of employer and employee as having the payoff matrix defined in Fig. 18.

**fMRI Scanning Parameters.** To optimize functional sensitivity in OFC we acquired the data using an oblique orientation of 30° to the anterior–posterior commissure line. A total of 580 volumes (19 min) were collected during the experiment in an interleave ascending manner. The imaging parameters were as follows: echo time, 30 ms; field of view, 192 mm; in-plane resolution and slice thickness, 3 mm; TR, 2 s. High-resolution T1-weighted structural scans (1 × 1 × 1 mm) were acquired for anatomical localization. Image analysis was performed by using SPM2 (Wellcome Department of Imaging Neuroscience, Institute of Neurology, London). Pre-processing included slice timing correction (centered at TR/2), motion correction, spatial normalization to a standard T2* template with a resampled voxel size of 3 mm, and spatial smoothing using an 8-mm Gaussian kernel. Intensity normalization and high-pass temporal filtering (128 s) were also applied to the data (34).

**Behavioral Data Analysis.** Details of the computational models used in the analysis are provided in SI Methods and summarized in Table 1. Model parameters were estimated by maximizing simultaneously the logistic log likelihood of the predicted decision probabilities generated by a model against the actual behavior of all subjects. Parameters for the employer and employee roles were fitted separately to account for any differences. We used the multivariate constrained minimization function (fmincon) of the Optimization Toolbox 2.2 in Matlab 6.5 (Mathworks) for this fitting procedure (see SI Methods and Table S1 for more details). All behavioral data shown correspond to the actions of all 30 participants.

**fMRI Data Analysis.** The event-related fMRI data were analyzed by constructing sets of delta (stick) functions at the time of the choice and at the time of the outcome. Additional regressors were constructed by using the model estimated choice expected values as a modulating parameter at the time of choice and one or more (depending on the model) update signals as modulating parameters at the time of outcome. All of these regressors were convolved with a canonical hemodynamic response function. In addition, the six scan-to-scan motion parameters produced during realignment were included to account for residual motion effects. These were fitted to each subject individually, and the regression parameters were then taken to the random-effects level. All reported fMRI statistics and uncorrected P values arose from group random-effects analyses (n = 15), and small-volume corrected P values were obtained from 8-mm spheres around regions of interest previously implicated in mentalizing from a metaanalysis by Frith and Friston (3). Specifically, coordinates were defined from the center of each of the areas delineated from the metaanalysis: mPFC (0, 56, 19), STS (±53, −51, 10), and the temporal poles (±46, 11, −35 mm). An across-subject model likelihood difference modulator was also fitted at the second level, alongside the mean group level effect, when studying the influence signal in Fig. 3.

We also compared the influence and RL algorithms with each other by fitting both models at the same time against the fMRI data (Fig. 2B). To make the models as similar as possible, we normalized all modulating regressors before fitting to the fMRI data. Areas showing significant activity for any regressor of a given model indicated regions that were better explained by that model in comparison to the other. This approach for model comparison, as well as the alternative ROI approach, is explained in more detail in SI Methods.

**ACKNOWLEDGMENTS.** J.P.O. was supported by grants from the Gimbel Discovery Fund in Neuroscience, the Gordon and Betty Moore Foundation, and a Searle Scholarship. P.B. was supported by the Swiss Finance Institute.

9. McCabe K, Houser D, Ryan L, Smith V, Trouard T (2001) A functional imaging study of payoff matrix defined in Fig. 1
Supporting Information

Hampton et al. 10.1073/pnas.0711099105

SI Methods

Computational Models. Reinforcement learning. Reinforcement learning (RL) is concerned with learning predictions of the future reward that will be obtained from being in a particular state of the world or performing a particular action. In this paper we use a simple RL model in which action values are updated via a Rescorla-Wagner (RW) rule (1). On a trial $t$ in which action $a$ is selected, the value of action $a$ is updated via a prediction error $\delta$:

$$ V^a_{t+1} = V^a_t + \eta \delta, $$

where $\eta$ is the learning rate. The prediction error $\delta$ is calculated by comparing the actual reward received $R_t$ after choosing action $a$ with the expected reward for that action:

$$ \delta = R_t - V^a_t. $$

When choosing between two different states ($a$ and $b$), the model compares the expected values to select which will give it the most reward in the future. The probability of choosing action $a$ is

$$ p^a = f(V^a - V^b), $$

where $f(z) = 1/(1 + e^{-z})$ is the Luce choice rule (2) or logistic sigmoid, and $\beta$ reflects the degree of stochasticity in making the choice (i.e., the exploration/exploitation parameter).

Fictitious play. In game theory, a first order fictitious play model (3) is one in which a player infers the probability that the opponent will choose one action or another, and then decides so as to maximize the action’s consequent expected reward. The opponent’s probability $p^{**}$ of choosing action $a^*$ is dynamically inferred by tracking the history of actions the opponent makes:

$$ p^{**}_t = p^*_t + \eta \delta^p, $$

where $\delta^p = P_t - p^*_t$ is the prediction error between the opponent’s expected action $p^*$ and whether the opponent chose action $a^*$ at time $t$ ($P_t = 1$), or chose another action ($P_t = 0$). The opponent’s action probabilities $p^*$, the expected value for each of the player’s actions can be calculated using the payoff matrix of the game. A stochastic choice probability can then be calculated using Eq. 3. For the inspection game described in this paper, this can be summarized as follows: calling $p$ the probability that the employee will work, and $q$ the probability that the employer will not inspect, and using the payoff matrix of the game (Table 1—in the following formulations, payoffs were express in 25 cent units for convenience), the decision of each player is

$$ p = f(2 - 4q), $$

$$ q = f(5p - 1), $$

where $q^*$ and $p^*$ are the inferred probabilities of the opponent’s actions estimated using Eq. 4.

An equivalent formulation of fictitious play is one in which the values of actions are learned directly as in reinforcement models, instead of tracking the opponent player’s action probability. For this, not only the value of the chosen action is updated with the reward that was received (as in Eq. 1), but also all other actions are penalized proportional to their foregone rewards (4, 5). Either approach posits knowledge of the structure of the game to update the variable estimates and arrive at a correct expected value for the actions of each player.

Experience Weighted Attraction (EWA): The EWA learning rule we used here is a combination of Reinforcement Learning and Fictitious play. It updates the value of a choice such that:

$$ V_{t+1}^a = (1 - \eta)V_t^a + \eta(\delta R_t + (1 - \delta)R_2), $$

where $R_1$ is the reward obtained had action $a$ been chosen (Fictitious learning), and $R_2$ is the reward given that action $a$ was chosen — zero otherwise (Reinforcement Learning). Some variants of EWA also involve an additional parameter which modulates the rate of learning at different points in the game, such that it can be faster at the beginning of a game, and then become slower as subjects settle into a strategy towards the end. In this study we assumed constant learning throughout the game, so did not include this additional parameter.

Influencing the Opponent. How much does a player’s next decision change given the action of the opponent? Replacing the update of the inferred opponent’s strategy (Eq. 4) in a player’s decision (Eq. 5), and Taylor expanding (around $\eta = 0$),

$$ \Delta p = -\eta \beta p_t(1 - p_t)(q - q^*) $$

$$ \Delta q = +\eta \beta q_t(1 - q_t)(p - p^*). $$

The sign difference in both terms is determined by the competitive structure of the game; namely, that the employer wants to inspect when the employee shirks, while the employee wants to shirk when the employer does not inspect. A player can obtain a more accurate inference of the opponent’s action strategy by incorporating the influence his/her own action has on the opponent. Thus, at the end of each trial both players update the estimates of their opponent such that

$$ p_{t+1} = p_t + \eta_1(P_t - p_t^*) - \eta_2 \beta p_t(1 - p_t^*)(Q_t - q_t^*) $$

$$ q_{t+1} = q_t + \eta_1(Q_t - q_t^*) + \eta_2 \beta q_t(1 - q_t^*)(P_t - p_t^*); $$

where $q^{**}$ and $p^{**}$ are the inferred probabilities that the opponent has of the player itself (second-order beliefs). Thus, this gives two clear signals: the prediction error as the first term and the influence update as the second term. The influence update, or how much a player influences his/her opponent, is proportional to the difference between the action a actor took and what the opponent thought was the player’s strategy. These second order beliefs can be inferred by the player directly from the inferred opponent’s strategy by inverting Eq. 5.

$$ p_t^* = \frac{4}{3} + \frac{1}{5\beta} \log \left( \frac{1 - q_t^*}{q_t} \right) $$

$$ q_t^* = \frac{4}{5} + \frac{1}{4\beta} \log \left( \frac{1 - p_t^*}{p_t} \right). $$

Behavioral Data Analysis. The RL, Fictitious play, and Influence model decision probabilities $p$ (action $a$ or $b$)—the probability a certain action would be taken predicted by the model (Eq. 3)—were fitted against the behavioral data $y$ (action $a$ or $b$)—the actual behavioral choice made by the subject. The parameters of each model were fitted against all subjects responses by maximizing the logistic log likelihood of the model predictions.
with one set of parameters modeling the employer role, and another the employee role. We used the multivariate constrained minimization function (fmincon) of the Optimization Toolbox 2.2 in MATLAB 6.5 (MathWorks) to estimate the model parameters given the log likelihood defined above.

The fitted parameters for each model are shown in Table S1.

fMRI Model Comparison: Voxel-Based Analysis. To compare the explanatory power of signals predicted by two competing models, we fit both models simultaneously to the brain BOLD signals (Fig. 2B) and then test for significance of a particular regressor (in this case, a random effects t test of the Influence model’s expected reward signal). When testing a particular regressor for significance, SPM uses the extra sum of squares test by comparing the variance explained by a full model containing this regressor to a partial model not containing that regressor, e.g., (SS(Influence + RL) − SS(RL))/MSE; where MSE = residual error (6). In Fig. 2B, we report the additional variance explained by value signals from the Influence model above and beyond that explained by such signals from the RL model. It should be noted that because the free parameters for each model were fit on the behavioral data and not on the imaging data, the model comparison procedure on the imaging data are not affected by the different number of parameters in the model-fits to the behavioral data. Thus, there is no need to correct for different numbers of free parameters in the models (using for instance AIC methods) when performing the model-comparison procedure on the fMRI data.

fMRI Model Comparison: ROI Analysis. A second approach consists of fitting both models separately to brain activity in a region of interest, and then comparing their regression coefficients to determine which model provides a better account of neural activity in this region. To test how well the expected reward signals from each model predicted BOLD activity in mPFC we provided more information when predicting activity in mPFC, than did linear models which only contained signals from either STS or ventral striatum. For example, to determine whether adding STS as a regressor to a model that already contains ventral striatum provided statistically significant information was calculated. A random effects statistic was then repeated for each different time point within a trial to plot the change in correlation between regions with respect to time within a trial for all subjects scanned in this study.

We also investigated the degree to which the time series extracted from STS and ventral striatum, as described above, were a significant predictor of the time series extracted from mPFC. In particular, we were interested in whether a linear model containing signals from STS and ventral striatum provided more information when predicting activity in mPFC, than did linear models which only contained signals from either STS or ventral striatum. For example, to determine whether adding STS as a regressor to a model that already contains ventral striatum as a regressor was a better predictor of activity in mPFC than the model with ventral striatum alone, we compared the likelihood of the complete model (LSTS,vStriatum) with the likelihood of the incomplete model (L,vStriatum). The difference in log likelihoods (Δ = 2 log LSTS,vStriatum − 2 log LSTS,S,vStriatum) follows a χ² distribution with one degree of freedom. Thus, for each subject scanned (and each game session), the probability of whether adding information from STS provided statistically significant information was calculated. A random effects statistic across subjects was then calculated by converting these individual p-values to z-scores, and then performing a t test across subjects.

Interregion Correlation Analysis. To compute inter-region correlations as reported in Fig. 4, a representative time series was obtained for each region by extracting and averaging BOLD activity from 10-mm spheres centered in the group peak for expected value in mPFC (−3, 63, 15 mm; see Fig. 2A); in the group peak for reward prediction errors in ventral striatum (9, 6, −18 mm, and −9, 9, −18 mm; see Fig. S3); and in the group peak for influence error signals in STS (−57, −54, 0 mm, and 60, −54, 9 mm; see Fig. S4). A general linear model with one regressor for each trial (totaling 100 regressors) was then created by convoluting a canonical hemodynamic response function with a stick function centered at the time of trial onset. This model was then fitted to the time series from each region of interest. The regression coefficients thus represent the deconvoluted neural activity for each trial of the game at the time of trial onset. The deconvoluted activity from each region of interest was then used to calculate the correlation between regions of interest at that time point. The deconvolution and correlation process was then repeated for each different time point within a trial to plot the change in correlation between regions with respect to time within a trial for all subjects scanned in this study.

We also investigated the degree to which the time series extracted from STS and ventral striatum, as described above, were a significant predictor of the time series extracted from mPFC. In particular, we were interested in whether a linear model containing signals from STS and ventral striatum provided more information when predicting activity in mPFC, than did linear models which only contained signals from either STS or ventral striatum. For example, to determine whether adding STS as a regressor to a model that already contains ventral striatum as a regressor was a better predictor of activity in mPFC than the model with ventral striatum alone, we compared the likelihood of the complete model (LSTS,vStriatum) with the likelihood of the incomplete model (L,vStriatum). The difference in log likelihoods (Δ = 2 log LSTS,vStriatum − 2 log LSTS,S,vStriatum) follows a χ² distribution with one degree of freedom. Thus, for each subject scanned (and each game session), the probability of whether adding information from STS provided statistically significant information was calculated. A random effects statistic across subjects was then calculated by converting these individual p-values to z-scores, and then performing a t test across subjects.

Fig. S1. Out-of-sample model log likelihood. The out-of-sample model log likelihood controls for models having different number of free parameters when fitting to behavioral data. Models were trained with the first 70 trials for each subject and then tested on the last 30 trials to obtain an out-of-sample log likelihood. The influence model accounts for subjects’ behavior the best, with an out-of-sample log likelihood of 0.674 ± 0.004, followed by the fictitious play model with 0.685 ± 0.003 and the RL model with 0.687 ± 0.003.
Influence > RL in Temporal Pole and STS

Fig. S2. Model comparisons with respect to the processing of expected reward signals in the brain. The influence model expected reward signals that are not explained by (orthogonal to) the RL model expected reward signals also activate the right STS, including the right temporal pole at $P < 0.001$ uncorrected.
Fig. S3. Prediction error signals. The prediction error signals generated by the influence model were correlated with activity in the ventral striatum bilaterally (9, 6, −18 mm, z = 4.97; −9, 9, −18 mm, z = 4.73, both P < 0.05 whole-brain corrected), mPFC (−9, 57, 6 mm, z = 4.35), and paracingulate cortex (12, 36, 18 mm, z = 4.62). This lends support to the suggestion that mPFC is not only involved in calculating expected reward signals derived from inference of the opponent’s game strategy (Fig. 2A), but is also involved in the update of the inferred opponent’s strategy through prediction errors (this figure) and influence updates (Fig. 3B).
<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Employer</th>
<th>Employee</th>
</tr>
</thead>
<tbody>
<tr>
<td>RL</td>
<td>$\beta$</td>
<td>0.011</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td>$\eta$</td>
<td>0.16</td>
<td>0.11</td>
</tr>
<tr>
<td>Fictitious play</td>
<td>$\beta$</td>
<td>0.036</td>
<td>0.056</td>
</tr>
<tr>
<td></td>
<td>$\eta$</td>
<td>0.22</td>
<td>0.10</td>
</tr>
<tr>
<td>Influence model</td>
<td>$\beta$</td>
<td>0.090</td>
<td>0.059</td>
</tr>
<tr>
<td></td>
<td>$\eta$</td>
<td>0.21</td>
<td>0.038</td>
</tr>
<tr>
<td></td>
<td>$\kappa$</td>
<td>0.0011</td>
<td>0.043</td>
</tr>
</tbody>
</table>