Neural correlates of depth of strategic reasoning in medial prefrontal cortex

Giorgio Coricelli[a,b,1] and Rosemarie Nagel[c]

[a]Cognitive Neuroscience Centre, Centre National de la Recherche Scientifique, 67 Boulevard Pinel, 69675, Bron (Lyon), France; [b]Center for Mind/Brain Sciences, Centro interdipartimentale Mente/Cervello, University of Trento, Via delle Regole, 101, 38060 Mattarello, Trento, Italy; and [c]Institució Catalana de Recerca i Estudis Avançats, Department of Economics, Universitat Pompeu Fabra, Ramón Trias Fargas, 25–27 08005 Barcelona, Spain

Edited by Michael Gazzaniga, University of California, Santa Barbara, and accepted by the Editorial Board April 16, 2009 (received for review August 11, 2008)

We used functional MRI (fMRI) to investigate human mental processes in a competitive interactive setting—the "beauty contest" game. This game is well-suited for investigating whether and how a player's mental processing incorporates the thinking process of others in strategic reasoning. We apply a cognitive hierarchy model to classify subject's choices in the experimental game according to the degree of strategic reasoning so that we can identify the neural substrates of different levels of strategizing. According to this model, high-level reasoners expect the others to behave strategically, whereas low-level reasoners choose based on the expectation that others will choose randomly. The data show that high-level reasoning and a measure of strategic IQ (related to winning in the game) correlate with the neural activity in the medial prefrontal cortex, demonstrating its crucial role in successful mentalizing. This supports a cognitive hierarchy model of human brain and behavior.

bounded rationality | cognitive hierarchies | game theory | neuronimaging | theory of mind

"Professional investment may be likened to those newspaper competitions [Beauty Contest] in which the competitors have to pick out the 6 prettiest faces from a hundred photographs, the prize being awarded to the competitor whose choice most nearly corresponds to the average preferences of the competitors as a whole. It is not a case of choosing those which are really the prettiest, nor even those which average opinion genuinely thinks the prettiest. We have reached the third degree—to anticipating what average opinion expects the average opinion to be. And there are some, I believe, who practise the fourth, fifth, and higher degrees (1)."

John Maynard Keynes, one of the most influential economists of the 20th century, describes in the above quote different ways of thinking about others in a competitive environment. This can range from low-level reasoning, characterized by self-referential thinking (choosing what you like without considering others' behavior), to higher levels of reasoning, taking into account the thinking of others about others ("third degree"), and so on.

Many features of social and competitive interaction require this kind of reasoning, for example, deciding when to queue for precious theater tickets or when to sell or buy in the stock market before too many others do it. Psychologists and philosophers define this as theory of mind or mentalizing, the ability to think about others' thoughts and mental states to predict their intentions and actions (2–9). Neuroimaging studies have found brain activity related to mentalizing in the medial prefrontal cortex (3, 5, 6, 10–12), temporo-parietal junction (3, 13), superior temporal sulcus (14), and posterior cingulate cortex (5). However, little is known about the neural mechanisms underlying the iterated steps of thinking, "what you think the others think about what you think," and so on. That is, the mechanisms underlying how deeply people think about others, and, particularly, whether deeper mentalizing leads to more successful social outcomes.

Here, we study an experimental competitive game, analogous to Keynes's Beauty Contest, to characterize the neural systems that mediate different levels of strategic reasoning and mentalizing. In our experimental game, participants choose a number between 0 and 100. The winner is the person whose number is closest to 2/3 times the average of all chosen numbers (Fig. 1*A* and *Methods*).

Game theory suggests a process of iterated elimination of weakly dominated strategies, which in infinite steps reaches the unique Nash equilibrium in which everybody chooses 0 (Fig. 1*B*). However, "the natural way of looking at game situations is not based on circular concepts [as for the Nash equilibrium] but rather on a step by step reasoning procedure (ref. 15, p. 421)," which typically results in out-of-equilibrium behavior. This step-reasoning can be some finite steps of the iterated elimination process (Fig. 1*B*) or of the so-called iterated best reply, a cognitive hierarchy of thinking that better describes behavior in our game (Fig. 1*C*) (16–18). In our game, this means that a naïve player (level 0) chooses randomly. A level 1 player thinks of others as level 0 reasoning and chooses 33 (= 2/3*50), because 50 is the average of randomly chosen numbers from 0 to 100. A more sophisticated player (level 2) supposes that everybody thinks like a level 1 player and therefore he chooses 22 (= $(2/3)^2$ *50). And, as Keynes mentioned there might eventually be people reaching the (Nash) equilibrium of the game and thereby choosing 0. Choices in many Beauty Contest experimental games (17, 19–21), but also in other games (16, 18), show limited steps of reasoning, a bounded rational behavior, confirming the relevance of the iterated best-reply model (see *SI Text SI1*).

Why do people use different and limited numbers of steps of reasoning? As the number of steps of thinking increases, the decision rule requires more computation, and higher level reasoning indicates more strategic behavior paired with the belief that the other players are also more strategic (16). One reason for the limited steps of reasoning is that players might be incapable of using high levels of reasoning because of cognitive limitations (22); another reason is that a player might believe (overconfidently) (23) that others will not use as many steps of thinking as he does. Identifying the neural correlates of different levels of reasoning and, more specifically, being able to distinguish between low- versus high-level reasoning people by their brain activity will help to explain the heterogeneity observed in human strategic behavior.

We used functional MRI (fMRI) to measure brain activity when subjects participated in the Beauty Contest game. We introduced 2 main conditions in an event-related fashion (Fig. 1*A* and *Methods*). In the human condition, each participant in a group of 10 was asked to choose an integer between 0 and 100. In the computer condition one participant chose one number between 0 and 100 and a computer algorithm chose uniform randomly (and independently of the multiplier parameter) 9 numbers between 0 and 100. The

Author contributions: G.C. and R.N. designed research; G.C. performed research; G.C. and R.N. analyzed data; and G.C. and R.N. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. M.G. is a guest editor invited by the Editorial Board.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. E-mail: coricelli@isc.cnrs.fr.

This article contains supporting information online at www.pnas.org/cgi/content/full/0807721106/DCSupplemental.

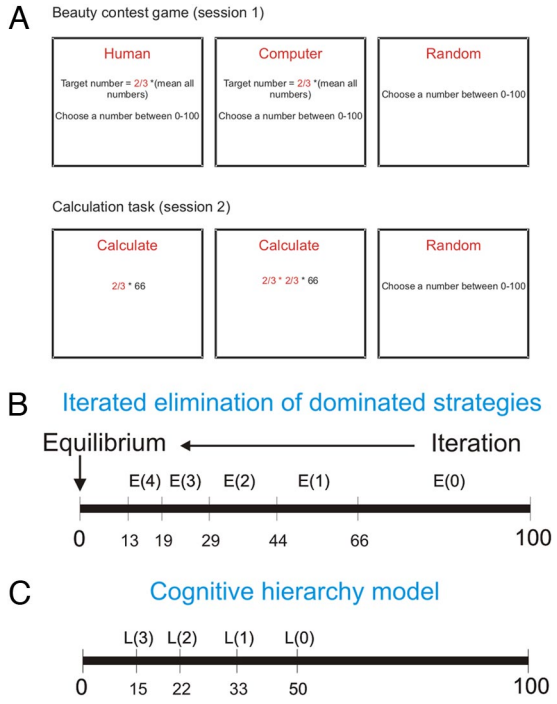

Fig. 1. (*A*) Rules of the basic game and conditions. The figure shows the computer screen for each experimental condition. The participants were asked to choose a number between 0 and 100. The winner got 10 euros (or an equal share with those who tie) and is the person whose number is closest to the target [a parameter multiplier (here 2/3) times the average of 10 numbers]. The 10 numbers are the choices of either 10 human participants (human condition) or of one participant and a computer program who chooses uniform randomly 9 numbers from 0 to 100 (computer condition). The losers got nothing. All this is known to the participants. There were 13 different parameter multipliers. Each multiplier was presented once in each condition in a pseudorandom order. In a control condition (random condition) the participants were asked to pronounce a random number between 0 and 100. In the calculation task (session 2) subjects were asked to calculate the product between one (C1 condition) or 2 factors (C2 condition) times a number, and additionally a random condition. (*B*) Game theoretic prediction (for M = 2/3): If all participants are rational and know that everybody else is rational and so on (common knowledge of rationality) then everybody should choose 0, because no one should choose $>100*2/3 = 66$ (weakly dominated choices); thus all numbers in E (0) are eliminated. In the reduced game nobody should choose $>100*(2/3)^2 = 44$, thus eliminating E (1), and so on until 0 is reached. If M > 1 then all players choosing 100 also represents an equilibrium. (*C*) Bounded rational model. Cognitive hierarchy (for M = 2/3) is a cognitively and descriptively more plausible model (17). A random player level 0 [L (0)] chooses uniformly from 0 to 100 with an average of 50. A best reply to this is $50*2/3 = 33$ (level 1). If everybody chooses 33 then best reply is $50*(2/3)^2 = 22$ (level 2), etc. A subject is strategic of degree k if he chooses the number $50*M^k$, called level k.

prize for the winner, whose number was closest to M (e.g., M = 2/3) times the average of all choices, was 10 euros in both conditions or a split of the prize in the case of a tie. We did not provide any feedback between trials. The computer condition should invoke low levels of reasoning (at or near level 1) according to the iterative reply model. In contrast, in the human condition a greater variety of levels of reasoning should be observed because players might have different ideas about what other players choose. To identify brain activity related to the mental calculation most likely used when deciding in the game, we introduced calculation tasks in which subjects were asked to multiply a given parameter or the square of a parameter with a given integer.

Results

Behavioral Results. Reaction time was quite different in the different conditions of the Beauty Contest. Subjects took longer when choosing a number in the human (mean = 8.94, SD = 7.6) compared with the computer condition (mean = 7.28, SD = 5.6; Wilcoxon signed-rank test, $z = 2.2$, $P = 0.03$, two-tailed). In both conditions, choosing took more time than in a control condition when they were asked to pick a random number between 0 and 100 (mean = 2.03, SD = 1.07) (for both human vs. random and computer vs. random signed-rank tests, $z = 3.92$, $P < 0.001$).

Bounded Rational Behavior: Participants Played According to the Cognitive Hierarchy Model. As found in previous experimental economics studies of the Beauty Contest game (17, 19–21) the behavioral results confirmed the presence of play according to step-by-step reasoning of the iterated best reply model $50*M^k$, where k is the number of levels (Fig. 1*C*), and very seldom by the game theoretic solution (0 for M < 1, and 0 or 100 for M > 1) (see *SI Text SI2*). Most choices in the human condition were between L1 ($50*M$) and L3 ($50*M^3$), only 5% were higher than level 3. We measured the level of reasoning using the quadratic distance between actual choices and the different theoretical values (L1, L2, L3, etc.) based on the Cognitive Hierarchy model (see *Methods*). We categorized each player according to 3 categories (based on choices in the human condition): random behavior and low level (level 1) and high level of reasoning (level 2 or higher). The subjects classified as low level ($n = 10$) behaved similarly against the computer or the humans, at or close to level 1 in both conditions (Wilcoxon signed-rank test of the mean quadratic distance between actual choices and theoretical L1 across all trials for each subject in human vs. computer condition, $z = 0.76$, $P = 0.44$). The high-level reasoning subjects ($n = 7$) differentiated their behavior in the human compared with the computer condition (signed-rank test of the mean quadratic distance between actual choices and theoretical L1 across all trials for each subject in human vs. computer condition, $z = 2.36$, $P = 0.018$). They behaved as level 1 in the computer condition but were classified at a higher level of reasoning (level 2 or more) when interacting with human counterparts. Direct comparison between the 2 groups confirms that high reasoners have a significantly smaller quadratic distance between their actual choices and the theoretical level 2 (or higher) compared with low reasoners (Two-sample Wilcoxon rank-sum (Mann–Whitney) test, $z = -3.22$, $P = 0.0013$) (see *SI Text SI3*). Three subjects behaved in a quite random fashion. Fig. 2*A* shows the behavior, separately for each condition and for all parameter values, of 2 representative subjects (individual behavior of all of the subjects is shown in Fig. S1). In the computer condition, both subjects chose numbers close to or on the level 1 line ($50*M$, were M is the multiplier parameter). In the human condition, the low-level subject typically chose near the level 1 line, whereas the high-level subject chose near or at the level 2 line ($50*M^2$) or near or at a higher level.

fMRI Data: Medial Prefrontal Cortex Dissociates Between High- and Low-Level Strategic Reasoning. We found enhanced brain activity in the medial prefrontal cortex (mPFC), rostral anterior cingulate cortex (rACC), superior temporal sulcus (STS), posterior cingulate cortex, and bilateral temporo-parietal junction (TPJ) when subjects made choices facing human opponents rather than a computer (as shown in Fig. S2, and Table S1, random effect analysis human vs. computer, $n = 20$). This suggests that playing against other subjects in the Beauty Contest game activated the mentalizing network (3, 5, 6, 24, 25).

When we analyzed high- and low-level reasoning subjects separately [region of interest (ROI) analysis, see *Methods*], we found the activity in the medial prefrontal cortex related to the contrast human vs. computer to be significant only for the subjects classified as high level (Fig. S2*B*). In the high-level reasoners ($n = 7$), choosing a number in the human condition in contrast to the computer condition activated 2 main regions of the medial prefrontal cortex (Fig. S2*B*): a more dorsal (peak MNI coordinates, x = 0, y = 48, z = 24; commonly related to third person perspective

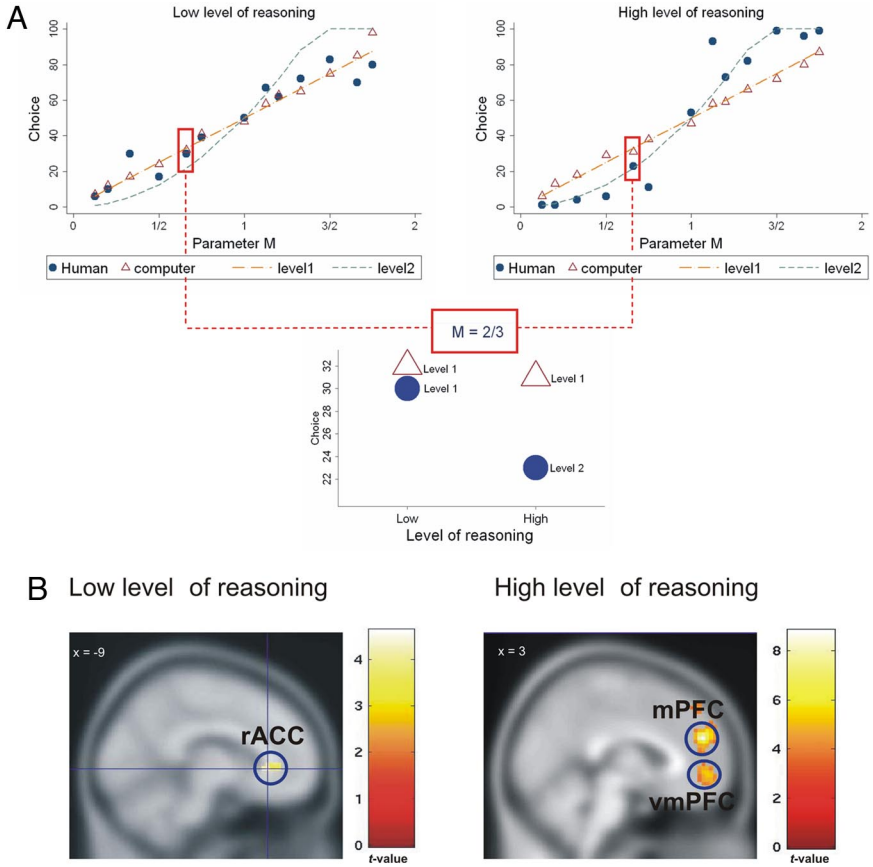

Fig. 2. Patterns of behavior and brain activity for low and high levels of reasoning. (*A*) Behavioral results. Here, we present the 26 choices of 2 representative participants for each parameter value M in the human (blue dots) and computer (triangles) conditions, separately. (*Left*) the choices of one participant representing a so-called low-level type. In both the computer condition (triangles) and the human condition (blue dots) he chose near the theoretical (Cognitive Hierarchy Model) level 1 line (brown line with choices equal to 50*M). (*Right*) the choices of one high-level type participant. In the computer condition he chose near the theoretical level 1 line. In the human condition he chose near the theoretical level 2 line (blue line with choices equal to $50*M^2$). (*Below*) We plot the choice of the 2 participants for the computer and human conditions for M = 2/3. In total we classified 10 participants as level 1 (low level) and 7 as greater than level 1 (high level). Three participants played in a random manner. (*B*) fMRI results. Choosing a number in the human condition in contrast to the computer condition was associated with relative enhanced activity in the rostral anterior cingulate cortex (*Left*) Low level of reasoning subjects, random effect analysis $n = 10$ (rACC, MNI coordinates: x = −9, y = 36, z = 3); and (*Right*) high level of reasoning subjects, random effect analysis $n = 7$ activity in the dorsal portion of the medial prefrontal cortex (mPFC, MNI coordinates: x = 3, y = 48, z = 24) and ventral mPFC (relatively less deactivated, MNI coordinates: x = 3, y = 51, z = −9). Group data thresholded at $P < 0.001$, uncorrected, is plotted on sagittal sections of a normalized canonical template brain.

taking) (6, 26) and a more ventral portion of the mPFC (relatively less deactivated in the human vs. computer, peak MNI coordinates, x = 9, y = 48, z = −6; related to thinking about a like minded or similar other) (27–29).

The main prefrontal activity of the low-level reasoning subjects ($n = 10$) was found in the rostral anterior cingulate cortex (rACC, peak MNI coordinates, x = −9, y = 36, z = 3, Fig. S2*B*), an area often attributed to self-referential thinking in social cognitive tasks (30–32). This suggests a dissociable contribution of different portions of the prefrontal cortex related to different behavioral responses (based on a different level of reasoning about others) in the Beauty Contest game.

Neural Substrates Related to Complex Cognitive Processes Correlated with High Level of Reasoning about Others. fMRI results (from the ROI analysis) show additional brain activities related to high- versus low-level reasoning in the right and left caudolateral orbitofrontal cortex (peak MNI coordinates x = 36, y = 24, z = −18 and x = −45, y = 21, z = −18, respectively), and left and right dorsolateral prefrontal cortex (peak MNI coordinates, x = −30, y = 18, z = 36 and x = 33, y = 15, z = 42, respectively), areas likely related to performance monitoring and cognitive control (33–35) (Fig. S3). This suggests that a complex cognitive process subserves the higher level of reasoning about others.

Ascription of Human-Human Interaction in Temporo-Parietal Junction and Superior Temporal Sulcus. The TPJ has been described as playing an important role in theory of mind tasks (13, 36, 37). Our data suggest that the activity of the TPJ (peak MNI coordinates, right TPJ: x = 51, y = −60, z = 27 and left TPJ: x = −51, y = −60, z = 21) is related to playing against humans independently of the level of reasoning (low or high). Similarly, activity in the superior temporal sulcus [left STS, peak MNI coordinates, x = −51, y = −18, z = 12], often related to the attribution of intentions and causality (38, 39) was found in both high- and low-level reasoning subjects (Fig. S4).

Fronto-Parietal Activity Related to Mental Calculation. The Beauty Contest game also requires solving a complex calculation task. Thus, to follow a first or higher level of reasoning, the subjects need to mentally think what might be the average of the numbers guessed by the others or chosen by the computer, including into this average their own number, and then multiplying the result by the announced factor, one or more times. Bilateral activity in the parietal cortex, encompassing the angular gyrus (peak MNI coordinates, x = 39,

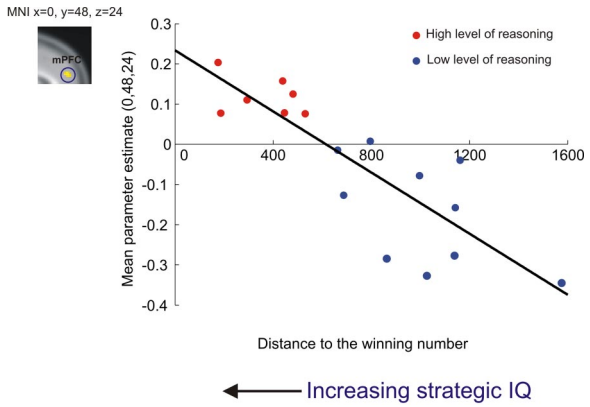

Fig. 3. Strategic IQ and medial prefrontal cortex. Activity in the dorsal portion of the medial prefrontal cortex related to play against human opponents (mPFC, MNI coordinates x = 0, y = 48, z = 24; the mean parameter estimates for each participant were extracted from the functional ROI obtained from the random effect analysis human vs. computer, $n = 20$) was correlated with a measure of strategic IQ (the quadratic distance of choices to the winning numbers using a recombinant estimation method). Values closer to 0 indicate higher strategic IQ. Note: red dots and blue dots indicate high and low level of reasoning participants, respectively; participants who played in a random manner are excluded from the figure.

y = −51, z = 24), the inferior parietal lobule, and the supramarginal gyrus (peak MNI coordinates, x = 54, y = −51, z = 30), was found both in the human and computer conditions of the Beauty Contest. Results from our calculation task show enhanced activity in the angular gyrus (peak MNI coordinates, x = 45, y = −42, z = 39) and in the inferior parietal lobule (peak MNI coordinates, x = 54 y = −42, z = 51) when the subjects were requested to mentally multiply a factor times a number (C1 condition), and when they were asked to multiply twice the same factor times a number (C2 condition). This suggests that part of the calculation activity related to the Beauty Contest game might be performed by these portions of the parietal cortex. Additional activity related to calculation (both C1 and C2 conditions) was found in the lateral prefrontal cortex (MNI coordinates, x = −48, y = 48, z = −3). This is consistent with previous findings on number processing and calculation (40–42). Importantly, no activity in the medial prefrontal cortex was related with any kind of calculation (Fig. S5).

Activity in the Medial Prefrontal Cortex Correlates with Strategic IQ. We found a cross-subject correlation (Pearson correlation $r = 0.67$, $P = 0.0013$) between the brain activity in the dorsal mPFC related to play against human opponents [MNI coordinates, x = 0, y = 48, z = 24; we obtained this functional ROI from the contrast human vs. computer, random effect analysis with $n = 20$; we then extracted from this ROI the beta values for each subject in the human condition] and a measure of strategic IQ in the Beauty Contest game [computed as the average of the quadratic distances of own choice to the winning numbers (M*average of 10 chosen numbers) across all trials and all possible combinations of the 19 opponents' choices; see also *Methods* section] (Fig. 3). Strategic IQ is reflected by the ability of subjects to match the right guess using higher levels of reasoning, that is, the ability to think deeply about others. It is therefore also a measure of economic earnings of a subject (43). Strategic IQ was not correlated (Pearson correlation $r = 0.09$, $P > 0.5$) with accuracy (number of correct responses) in the calculation task; thus it is independent of computation skills. Notably, no other brain region of interest was correlated with strategic IQ.

Discussion

In the experimental Beauty Contest game, levels of reasoning are not induced [unlike the tasks used by (26, 43)]. Therefore, we could detect heterogeneity between subjects based on their own choice of depth of reasoning. We provide a computational account (Cognitive Hierarchy Theory; refs. 16–18) of the cognitive processing underlying actual choices in the experimental game, to identify the neural substrates of different levels of strategic thinking.

We found that playing against human opponents versus a computer (programmed to play randomly) in the Beauty Contest game activated areas commonly associated with theory of mind or mentalizing–thinking about other people's minds (mPFC, STS, posterior cingulate cortex, and TPJ) (3, 5, 6, 24, 25), suggesting that these areas encode the complexity underlying human interactive situations. Within this network, the mPFC was the only area that clearly dissociated between subjects with different levels of strategic reasoning. The mPFC activity (peak MNI coordinates, x = 0, y = 48, z = 24) differed in the human versus computer opponent conditions for high reasoning players only (Fig. S2*B*). Furthermore, in the human condition, this area was more active for high than low reasoners. Thus, we argue that mPFC implements more strategic thinking about other players' thoughts and behavior.

We also found that, unlike the mPFC, TPJ and STS mediated activity when playing against humans for both low and high-level reasoners. This suggests that the TPJ and STS have a more general function in the recognition of social cues or in the ascription of generic features of human-human interaction (44).

An additional insight into the role of the mPFC in social-cognitive processing is provided by the analysis of our measure of Strategic IQ (related to winning in the game). We found a strong correlation between mPFC activity and Strategic IQ. This suggests that the mPFC activity, involved in higher reasoning about others, leads to successful outcomes in our social setting. This is a new finding in the theory of mind literature, thus providing evidence for the fundamental role of the mPFC in successful mentalizing.

Notably, the focus of activity in the mPFC (peak MNI coordinates, x = 0, y = 48, z = 24; related to higher level of reasoning in our game) coincides with the focus of activity related to degree of thinking about how our own behavior can influence others' behavior, as reported in a recent study (45). In the study by Hampton et al. (45) activity in the mPFC was found when contrasting 2 dynamic models of choice in a repeated competitive game. One model is based on updating own strategy based on other's past choices and giving best response to the frequency play of actual behavior. A second, more sophisticated model assumes that subjects consider the influence that their own past choices will have on what other players will do next. The difference is analogous to the difference in the Beauty Contest game between high and low levels of strategic reasoning. Indeed, high-level reasoning in the Beauty Contest game implies thinking about how other players think about the others' (including your own) thinking and behavior, and so on. In other words, high reasoners might assume that their behavior likely affects the behavior of others, thus inducing a process of iterative thinking.

Thus, we argue that mPFC encoding of the effect of our choices on others' thoughts and behavior is the neural signature of high-level strategic reasoning (level 2 or more). The main difference between Hampton et al. (45) and our study is that in Hampton's study subjects observed others' behavior over time and then responded to it, whereas in our study the decisions required that subjects model and predict others' choices without knowing other players' past choices. The brain does not seem to distinguish between these 2 data sources. Taken together, the results of these 2 studies represent the first neural evidence of a close link between adaptive learning and levels of reasoning.

The pattern of brain activity (that is, higher activity for high-reasoning players) in the caudolateral orbitofrontal cortex and in the dorsolateral prefrontal cortex, areas commonly associated with complex cognitive processing (33–35), together with the mPFC, suggests a substantial jump in complexity (beyond the mere calculation required by the decision rule of the Beauty Contest game, as

suggested by the fact that there was no activity in these areas related to the mental calculation in the control tasks, C1 and C2) when going from the first to the second level of reasoning. This might be responsible for the observed limited step-level reasoning, either because subjects are not able to make this jump or because they believe that not everybody else is able to make this jump.

Game theory predicts equilibrium play, assuming common knowledge of rationality—everybody is rational and thinks that everybody else is rational and so on. However, actual behavior deviates from equilibrium and is heterogeneous given different beliefs about others. Our work shows that the common tendency for humans to use boundedly rational strategies (cognitive hierarchies) is reflected in specialized neural substrates, such as the medial prefrontal cortex.

Methods

Subjects. Twenty healthy right-handed subjects (11 females) were recruited to take part in a study at the Neuroimaging Center of the Insitut des Sciences Cognitives (Bron, Lyon, France). Volunteers gave fully informed consent for the project which was approved by the French National Ethical Committee (Comité Consultatif de Protection des Personnes dans la Recherche Biomédicale). Each participant was screened to exclude medication and conditions including psychological or physical illness or history of head injury. Mean age of participants was 26 years (± 4.17, SD).

Experimental Design and Task. Each participant underwent fMRI scanning while performing a total of 99 trials of the experimental tasks (first session of 26 trials of the Beauty Contest game plus 13 random choices and a second session of 60 trials of a mental calculation task including 12 random choices). During scanning, the subject viewed a projection of a computer screen (see Fig. 1*A*) and gave a spoken response in each trial. The Beauty Contest game (session 1) consisted of guessing an integer number between 0 and 100 (both limits included), in which the winner is the person whose number is closest to M*(average of all chosen numbers). M is the known multiplier parameter in a trial which takes 6 values with $M < 1$ (1/8, 1/5, 1/3, 1/2, 2/3, 3/4), and 6 values with $M > 1$ (9/8, 6/5, 4/3, 3/2, 5/3, 7/4). We also include $M = 1$, which is a control, whether the thought process started at or ≈50. However, level 1 and level 2 cannot be distinguished for $M = 1$. The winner received 10 euros. If there was a tie, the 10 euros were split between those who tied. We did not provide any feedback between trials. Information about the results of the game was provided at the payment stage (see below). The first session consisted of 3 different conditions: (*i*) human condition, in which a subject knew that he was playing against 9 other subjects who were under exactly the same conditions as himself but at a different scheduled time (13 trials); (*ii*) computer condition, in which the subject was informed that a computer program randomly draws 9 numbers (13 trials); and (*iii*) a random condition, in which the subject was asked to choose a number at random between 0 and 100 (13 trials). Each value of the parameter M was presented twice, once in the human and a second time in the computer condition. We used an event-related design, mixing the 3 main conditions. The calculation tasks (session 2) were of the form N*M (24 trials) or N*M*M (24 trials), where N is a 2 digit number and M is a multiplier from the set mentioned above, excluding $M = 1$. We did not provide any feedback between trials. Each product was mentioned twice with the same M but different N. We also asked for a random number as a control (12 trials). For each correct calculation task a subject received 50 euro cents. A correct answer had to be within a +/− 1 deviation of the up or down rounded result, e.g., a result of 54.33 produced a winning interval from 53 to 56. The calculation task was always presented after the Beauty Contest game to avoid behavioral biases. The participants were informed and instructed about the calculation task just before it began.

Time Course of the Experimental Tasks. On each trial of the Beauty Contest game (session 1) the subject viewed an information screen (2 seconds) indicating the type of condition (human, computer, or random), the formula of the target number with the information about the value of the parameter multiplier M (with the exception of the random condition), and the question to choose a number between 0 and 100. After 2 seconds, the message ''press the button when ready'' appeared in the bottom of the screen. After pressing the button the subjects had 2 seconds to ''say a number.'' There was an intertrial interval of 4–8 s (jittered). In the calculation task (session 2) the subjects viewed an information screen (2 seconds) with the indication of the factor(s) and the number digit they had to multiply. They were asked to give an answer with a maximum of 10 s. The message ''say a number'' appeared right after they pressed the button, or automatically after the time limit (10 s). There was an intertrial interval of 4–8 s (jittered).

Stimuli Presentation. Behavioral responses were logged by means of a desktop computer located outside the scanner running Presentation (Neurobehavioral Systems, Inc.) (stimulus delivery and experimental control software system for neuroscience).

Questionnaire. At the end of the scanning subjects had to fill in a questionnaire with the following questions. (*i*) Please comment on your first choice. $M = 2/3$ in the human condition. (*ii*) Please comment on your choice when $M = 1/4$ in the computer condition. (*iii*) Did you have a general rule for the trials in the human condition? (*iv*) Did you have a general rule for the trials in the computer condition? Notably, the subjects' responses on the questionnaire were very consistent with their pattern of choice (see *SI Text SI4*).

Payments. Participants were financially motivated. Each subject received a 50 euro show-up fee at the end of the experiment. Once 10 subjects had been scanned, we sent them an e-mail message with a table of their own choices and the choices of their coplayers (preserving anonymity). For each subject, we summed up the winning amount of each trial according to the rules of the game mentioned above. A transfer of the money they won was sent to their bank account.

Statistical Analysis of Behavioral Data: Behavioral Types. We categorized each player according to 3 categories (based on choices in the human condition): random behavior, low level (level 1), and high-level reasoning (level 2 or higher). To measure the level of reasoning of a subject we did the following: (*i*) we calculated for each trial of the Beauty Contest game the quadratic distance (QD) between actual choice and the different theoretical level *k* values [L1, L2, L3, …, according to the Cognitive Hierarchy model (Fig. 1*C*)]:

$$QDM_{ij}k = (x_{\mathrm{i\,j\,M}} - 50{*}\mathrm{M}^k)^2 \qquad [1]$$

where x_{ijM} is the choice of participant i in condition j (either human or computer) in trial with parameter M; *k* is level *k* with $k = 1, 2, 3, \ldots$. (*ii*) We determined the minimum distance and the corresponding level *k* within each trial. For example, a choice of 24 for $M = 2/3$, has is minimum QD for $k = 2$ [i.e., $QD1 = (24 - 50{*}(2/3))^2 = 87$, $QD2 = (24 - 50{*}(2/3)^2)^2 = 4$, $QD3 = (24 - 50{*}(2/3)^3)^2 = 84$, …], thus we classified this choice as level 2. (*iii*) For the human condition, we counted how many times (out of 12 M-parameters, thus without $M = 1$, which is not predictive of level of reasoning) a player was identified by one of the above-mentioned levels k of reasoning. (*iv*) We categorized a player as low level if at least 7 of 12 cases (in the human condition) were identified as level 1. We categorized a player as high level if at least 7 of 12 cases (in the human condition) were identified as level 2 or higher. Players that did not belong to any of these categories were classified as random players. We used the mean quadratic distance between actual choices and one theoretical type (e.g., L1 or L2) across all trials for each subject to test for behavioral differences between conditions (human vs. computer) and types (high vs. low), as reported in the results section.

Strategic IQ. We define strategic IQ as the subject's ability to guess a number that could potentially win against a large population of opponents. We considered all of the possible combinations of 9 choices out of all 19 opponents' choices (human condition) a player can be matched with (we followed the recombinant estimation method) (46, 47). For each subject, we calculated: (*i*) the winning number for each combination of choices (including the considered subject's choice) per trial; (*ii*) the quadratic distance of a subject's choice to the winning number of every combination per trial; (*iii*) the average quadratic distance across all of the possible combinations and all trials of a subject (plotted in Fig. 3, *x* axis).

fMRI: Data Acquisition, Preprocessing and Statistical Analysis. Subjects were scanned using a 1.5T MRI scanner (Siemens Magnetom Sonata with an 8 channel head coil) performing the experimental task over 2 sessions. T2-weighted echoplanar images, optimized for blood oxygenation level-dependent (BOLD) contrast, were acquired. Each volume comprised 26 slices acquired continuously over 2.5 s (TE: 60 ms; interleaved acquisition; slice thickness 4 mm; 0.4 mm noncontiguous; parallel to the subject's anterior–posterior axis; in plane resolution: 3.44*3.44 mm^2; matrix size: 64*64), allowing for complete brain coverage. Additionally, a T1-weighted image was acquired at the end of each experiment. Head motions were minimized by the use of foam padding. Headphones and ear-plugs were used to dampen the scanner noise. We used an MRI-compatible microphone for recording voice responses.

Image preprocessing and subsequent analyses were done using statistical parametric mapping (SPM5; http://www.fil.ion.ucl.uk/~spm/SPM5.html) on a

Matlab platform. Images were initially realigned to correct for motion artifacts. Differences in the timing of images slices across each individual volume were corrected, and each volume was transformed into standard stereotaxic space and smoothed with a Gaussian filter (full-width half-maximum 8 mm). Voxel-wise differences in BOLD contrast within the smoothed normalized images resulting from the different task conditions and trial types were examined using SPM. Standard neuroimaging methods using the general linear model were used with the first level (individual subject analyses) providing contrasts for group effects analyzed at the second level. Choice-related neural activity at the time of choice was studied during the epoch between trial onset and subject response (self-paced button-press before pronouncing a number). The intertrial intervals were jittered using an optimal signal-to-noise function (48, 49).

Choice trials were partitioned according to whether the subject was in the human, computer, or random condition of the Beauty Contest game and in the C1, C2, or random condition in the calculation task. For group analysis of choice-related activity, second-level analyses of contrast for different levels of reasoning for different trial types (human, computer, and random) where computed as ANOVAs with sphericity correction for repeated measures. Posthoc exploration of individual data is also reported to illustrate specific effects as a function of different trial types. Adjusted activity represents BOLD signal changes proportionally adjusted for the analytic model. Although general threshold significance was set at $P < 0.05$, corrected for family-wise errors, we tabulated group effects at $P < 0.0001$ or $P < 0.001$, uncorrected, to highlight regions of interest ROI. Volumes of interest analyses were performed using the MarsBaR toolbox for SPM. We extracted the estimated regressor coefficients (the beta values) averaged across the voxels of functional clusters (from the random effect analysis with $n = 20$) for the main contrasts of interest (e.g., we plotted averaged beta values extracted from the contrast human vs. computer for high and low-level reasoning subjects, plotted in Figs. S2–S4). The significance of the difference in brain activity (in the regions of interest) between high- and low-level reasoning participants was estimated with the Mann–Whitney test (nonparametric test). The activity described in Fig. 2*B* refers to random effect analyses restricted to the low ($n = 10$, *Left*) and high ($n = 7$, *Right*) reasoning subjects.

ACKNOWLEDGMENTS. We are grateful to Karen Reilly, Nadège Bault, Bill Harbaugh, Philippe Domenech, Andrea Brovelli, and Antoni Bosch for their comments on an early version of the manuscript, Aniol Llorente for computational support, and Danielle Ibarrola and the staff at CERMEP for technical support. This work was supported by the Human Frontier Science Program Grant RGP 56/2005 (to G.C. and R.N.); Agence Nationale de la Recherche, France and Provincia Autonoma di Trento (G.C.); CREA Grant SEJ2006[hyhen]135; Spanish Ministry of Education; and Agència de Gestió d'Ajuts Universitaris i de Recerca, Generalitat de Catalunya (R.N.).

1. Keynes JM (1936) *The General Theory of Employment Interest and Money*. (Macmillan; London).
2. Dennett DC (1989) *The Intentional Stance* (MIT Press, Cambridge, MA).
3. Fletcher PC, et al. (1995) Other minds in the brain: A functional imaging study of ''theory of mind'' in story comprehension. *Cognition* 57:109–128.
4. Gallese V, Goldman A (1998) Mirror neurons and the simulation theory of mind-reading. *Trends Cogn Sci* 2:493–501.
5. Gallagher HL, et al. (2000) Reading the mind in cartoons and stories: An fMRI study of ''theory of mind'' in verbal and nonverbal tasks. *Neuropsychologia* 38:11–21.
6. Amodio DM, Frith CD (2006) Meeting of minds: The medial frontal cortex and social cognition. *Nat Rev Neurosci* 7:268–277.
7. Goldman A (2006) *Simulating Minds*. (Oxford Univ Press, New York).
8. Frith CD (2007) The social brain? *Philos Trans R Soc Lond B* 362:671–678.
9. Coricelli G (2005) Two-levels of mental states attribution: From automaticity to voluntariness. *Neuropsychologia* 43:294–300.
10. Frith CD, Frith U (2006) The neural basis of mentalizing. *Neuron* 50:531–534.
11. Gallagher HL, Jack AI, Roepstorff A, Frith CD (2002) Imaging the intentional stance in a competitive game. *NeuroImage* 16:814–821.
12. Walter H, et al. (2004) Understanding intentions in social interaction: The role of the anterior paracingulate cortex. *J Cognit Neurosci* 16:1854–1863.
13. Saxe R, Wexler A (2005) Making sense of another mind: The role of the right temporo-parietal junction. *Neuropsychologia* 43:1391–1399.
14. Saxe R, Xiao DK, Kovacs G, Perrett DI, Kanwisher N (2004) A region of right posterior superior temporal sulcus responds to observed intentional actions. *Neuropsychologia* 42:1435–1446.
15. Selten R (1998) Features of experimentally observed bounded rationality. *Eur Econ Rev* 42:413–436.
16. Camerer CF, Ho T-H, Chong J-K (2004) A cognitive hierarchy model of games. *Q J Econ* 119:861–898.
17. Nagel R (1995) Unraveling in guessing games: An experimental study. *Amer Econ Rev* 85:1313–1326.
18. Stahl DO, Wilson PW (1995) On players' models of other players: Theory and experimental evidence. *Games Econ Behav* 10:218–254.
19. Bosch-Domenech A, Montalvo JG, Nagel R, Satorra A (2002) One, two, (three), infinity: Newspaper and lab Beauty-Contest experiments. *Amer Econ Rev* 92:1687–1701.
20. Costa-Gomes M, Crawford VP (2006) Cognition and behavior in two-person guessing games: An experimental study. *Am Econ Rev* 96:1737–1768.
21. Ho T-H, Camerer CF, Weigelt K (1998) Iterated dominance and iterated best response in experimental ''p-Beauty contests''. *Am Econ Rev* 88:947–969.
22. Devetag G, Warglien M (2003) Games and phone numbers: Do short-term memory bounds affect strategic behavior? *J Econ Psychol* 24:189–202.
23. Camerer CF, Lovallo D (1999) Overconfidence and excess entry: Experimental evidence. *Am Econ Rev* 89:306–318.
24. Bird CM, Castelli F, Malik O, Frith U, Husain M (2004) The impact of extensive medial frontal lobe damage on ''Theory of Mind'' and cognition. *Brain* 127:914–928.
25. McCabe K, Houser D, Ryan L, Smith V, Trouard T (2001) A functional imaging study of cooperation in two-person reciprocal exchange. *Proc Natl Acad Sci USA* 98:11832–11835.
26. D'Argembeau A, et al. (2007) Distinct regions of the medial prefrontal cortex are associated with self-referential processing and perspective taking. *J Cognit Neurosci* 19:935–944.
27. Mitchell JP, Banaji MR, Macrae CN (2005) The link between social cognition and self-referential thought in the medial prefrontal cortex. *J Cognit Neurosci* 17:1306–1315.
28. Mitchell JP, Macrae CN, Banaji MR (2006) Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron* 50:655–663.
29. Jenkins AC, Macrae CN, Mitchell JP (2008) Repetition suppression of ventromedial prefrontal activity during judgments of self and others. *Proc Natl Acad Sci USA* 105:4507–4512.
30. Kelley WM, et al. (2002) Finding the self? An event-related fMRI study. *J Cognit Neurosci* 14:785–794.
31. Johnson SC, et al. (2002) Neural correlates of self-reflection. *Brain* 125:1808–1814.
32. Moran JM, Macrae CN, Heatherton TF, Wyland CL, Kelley WM (2006) Neuroanatomical evidence for distinct cognitive and affective components of self. *J Cognit Neurosci* 18:1586–1594.
33. Ridderinkhof KR, Ullsperger M, Crone EA, Nieuwenhuis S (2004) The role of the medial frontal cortex in cognitive control. *Science* 306:443–447.
34. Kerns JG (2006) Anterior cingulate and prefrontal cortex activity in an fMRI study of trial-to-trial adjustments on the Simon task. *NeuroImage* 33:399–405.
35. Koechlin E, Summerfield C (2007) An information theoretical approach to prefrontal executive function. *Trends Cogn Sci* 11:229–235.
36. Saxe R, Kanwisher N (2003) People thinking about thinking people. The role of the temporo-parietal junction in ''theory of mind''. *NeuroImage* 19:1835–1842.
37. Decety J, Lamm C (2007) The role of the right temporoparietal junction in social interaction: How low-level computational processes contribute to meta-cognition. *Neuroscientist* 13:580–593.
38. Brunet E, Sarfati Y, Hardy-Bayle MC, Decety J (2000) A PET investigation of the attribution of intentions with a nonverbal task. *NeuroImage* 11:157–166.
39. Castelli F, Happe F, Frith U, Frith C (2000) Movement and mind: A functional imaging study of perception and interpretation of complex intentional movement patterns. *NeuroImage* 12:314–325.
40. Gruber O, Indefrey P, Steinmetz H, Kleinschmidt A (2001) Dissociating neural correlates of cognitive components in mental calculation. *Cereb Cortex* 11:350–359.
41. Hubbard EM, Piazza M, Pinel P, Dehaene S (2005) Interactions between number and space in parietal cortex. *Nat Rev Neurosci* 6:435–448.
42. Piazza M, Pinel P, Le Bihan D, Dehaene S (2007) A magnitude code common to numerosities and number symbols in human intraparietal cortex. *Neuron* 53:293–305.
43. Bhatt M, Camerer CF (2005) Self-referential thinking and equilibrium as states of mind in games: fMRI evidence. *Games Econ Behav* 52:424–459.
44. Mitchell JP (2008) Activity in right temporo-parietal junction is not selective for theory-of-mind. *Cereb Cortex* 18:262–271.
45. Hampton AN, Bossaerts P, O'Doherty JP (2008) Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proc Natl Acad Sci USA* 105:6741–6746.
46. Mitzkewitz M, Nagel R (1993) Experimental results on ultimatum games with incomplete information. *Int J Game Theory* 22:171–198.
47. Mullin CH, Reiley DH (2006) Recombinant estimation for normal-form games, with applications to auctions and bargaining. *Games Econ Behav* 54:159–182.
48. Friston KJ, Zarahn E, Josephs O, Henson RN, Dale AM (1999) Stochastic designs in event-related fMRI. *NeuroImage* 10:607–619.
49. Liu TT, Frank LR, Wong EC, Buxton RB (2001) Detection power, estimation efficiency, and predictability in event-related fMRI. *NeuroImage* 13:759–773.