The neural basis of bounded rational behavior

Giorgio Coricelli[1,2] and Rosemarie Nagel[3]

Abstract

Bounded rational behaviour is commonly observed in experimental games and in real life situations. Neuroeconomics can help to understand the mental processing underlying bounded rationality and out-of-equilibrium behaviour. Here we report results from recent studies on the neural basis of limited steps of reasoning in a competitive setting – the beauty contest game. We use functional magnetic resonance imaging (fMRI) to study the neural correlates of human mental processes in strategic games. 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. We found a correlation between levels of strategic reasoning and activity in a neural network related to mentalizing, i.e. the ability to think about other's thoughts and mental states. Moreover, brain data showed how complex cognitive processes subserve the higher level of reasoning about others. We describe how a cognitive hierarchy model fits both behavioural and brain data.

Keywords: Game theory, Bounded rationality, Neuroeconomics

1. Introduction

Economists only recently departed from the rational man and the notion of common knowledge of rationality when theorizing on economic problems. Common knowledge of rationality means that a decision maker knows that he is rational, that he knows that the other decision makers are rational and that he knows that others also know that

[1] Department of Economics, University of Southern California, 3620 S Vermont Avenue, Los Angeles, CA USA, e-mail : giorgio.coricelli@usc.edu
[2] Institut des Sciences Cognitives, Centre of Cognitive Neuroscience, CNRS, 67 Blv. Pinel, 69675, Bron (Lyon), France.
[3] *ICREA, Department of Economics*, Universitat Pompeu Fabra, *Ramón Trias Fargas, 25-27 08005 Barcelona, Spain*. e-mail: rosemarie.nagel@upf.edu

everybody is rational, and so on. A rational agent maximizes his expected utility, which means that the utilities from different results are weighted by their objective or subjective probabilities and maximized. In the last two decades, experimental economists have provided experimental results showing how far humans comply with or deviate from these assumptions, thus corroborating theories of bounded rationality.

Here we use a neuroeconomics approach, combining economics and neuroscience, to study bounded rational behaviour determined by limited depth of reasoning on players' beliefs about one another in a competitive interactive setting – the beauty contest game. The game was inspired by a quote from Keynes (1936):

"Professional investment may be likened to those newspaper competitions [the beauty contest] in which the competitors have to pick out the six 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; so that each competitor has to pick, not those faces which he himself finds prettiest, but those which he thinks likeliest to catch the fancy of the other competitors, all of whom are looking at the problem from the same point of view. It is not a case of choosing those which, to the best of one's judgment, are really the prettiest, nor even those which average opinion genuinely thinks the prettiest. We have reached the third degree where we devote our intelligences to anticipating what average opinion expects the average opinion to be. And there are some, I believe, who practice the fourth, fifth and higher degrees." (Keynes, 1936).

Keynes describes 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' behaviour), to higher levels of reasoning, taking into account the thinking of others about others ("third degree"), and

so on. Note, however, that Keynes advises not to use either level 0 or level 1 ("*It is not a case of choosing those which, to the best of one's judgment, are really the prettiest, nor even those which average opinion genuinely thinks the prettiest.")*. However, he does not make a clear proposal what other level to choose.

Many features of social and competitive interaction require this kind of reasoning; for example, deciding when to queue for precious theatre tickets or when to sell or buy in the stock market, before too many others do it.

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. A player's tendency to believe that others will not use as many steps of thinking as he does might be due to cognitive limitations or individual characteristics, such as overconfidence (Camerer and Lovallo, 1999). A higher level of reasoning indicates more strategic behaviour paired with the belief that the other players are also more strategic (Camerer, Ho, and Chong, 2004).

Identifying the neural correlates of different levels of reasoning, and more specifically, being able to distinguish between low- *versus* high-level reasoning people according to their brain activity will help to explain the heterogeneity observed in human strategic behaviour.

2. The experimental beauty contest game

Nagel (1995) studies an experimental competitive game, analogous to Keynes's Beauty Contest, to characterise different levels of strategic reasoning. In the 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. This game is suitable for investigating whether and how a player's mental process incorporates the behaviour of

the other players in his strategic reasoning. 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**.**

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" (Selten, 1998, pp. 421) which typically results in out-of-equilibrium behaviour.

2.1 The cognitive hierarchy model

This step reasoning can be some finite steps of the iterated elimination process or of the so-called iterated best reply, a Cognitive Hierarchy of thinking, that better describes behaviour in the beauty contest game (Nagel, 1995; Stahl and Wilson, 1995; Camerer et al., 2004). For instance, a naïve player (level 0) chooses randomly. A level 1 player thinks of others as level 0 reasoning and chooses 33 (= 2/3*50), where 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. According to the Cognitive Hierarchy model a subject is strategic of degree k if he chooses the number $50*M^k$, called iteration step k. Choices in many beauty contest experimental games (Nagel, 1995; Ho, Camerer and Weigelt, 1998; Bosch-Domenech, Montalvo, Nagel, and Satorra, 2002; Costa-Gomes and Crawford, 2006) show limited steps of reasoning, a bounded rational behaviour, confirming the relevance of the iterated best-reply model. The Cognitive Hierarchy model: (1) is not an equilibrium model, i.e. strategies of players don't have to be best reply to each other; (2) it does not assume common knowledge of rationality; (3) it assumes that players best reply to own beliefs, which might be non consistent; (4) it is based on limited level of reasoning of oneself or others.

3. An fMRI study on levels of strategic reasoning

In Coricelli and Nagel (2009) we used functional magnetic resonance imaging (fMRI) to measure brain activity when subjects participated in the beauty contest game. We introduced two main conditions in an event-related fashion. In the *human condition*, each participant of a group of 10 was asked to choose an integer between 0 and 100. The winner is the person whose number is closest to the target number (a parameter multiplier (e.g., 2/3) times the average of the 10 chosen numbers within the group). In the *computer condition* one participant chose one number and a computer algorithm chose randomly (and independently of the multiplier parameter) nine numbers. This algorithm was known to the subjects. The prize for the winner was 10 euros in each trial of both conditions, or a split of the prize in case of ties. 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 higher variety of levels of reasoning should be observed since players might have different ideas what other players choose. To be able to identify brain activity related to mental calculation most likely involved when deciding in the game, we introduced calculation tasks in which subjects were asked to multiply a given parameter (e.g. 2/3*66) (C1 condition) or the square of a parameter (e.g., 2/3*2/3*66) (C2 condition) with a given integer.

3.1 Bounded rational behaviour: participants played according to the cognitive hierarchy model

As found in previous experimental economics studies of the game (e.g. Nagel, 1995; Stahl and Wilson, 1995; Bosch-Domenech et al., 2002; Camerer et al., 2004, Costa-Gomes and Crawford, 2006), in Coricelli and Nagel (2009) the behavioural results confirmed the presence of play according to the iterated best reply model. The starting point for the reasoning process was 50 and not 100, and the process was driven by ‘finite’

iterative best replies and not by elimination of dominated strategies. In the computer condition, all subjects chose numbers close to level 1 (50*M, were M is the multiplier parameter). We categorized each player according to three categories: random behaviour, level 1, and level 2 or higher reasoning. We measured the level of reasoning of a subject as the smallest quadratic distance between actual play and the different theoretical values based on the Cognitive Hierarchy model in the human condition. The high-level reasoning subjects (N=7) clearly differentiated their behaviour in the human compared to the computer condition. They behaved as level 1 in the computer condition but were classified as higher level of reasoning (level 2 or more) when interacting with human counterparts. 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. Three subjects behaved in a quite random fashion.

3.2. Neural correlates of depth of reasoning

In our fMRI study we found enhanced brain activity in the medial prefrontal cortex (mPFC), rostral anterior cingulate (ACC), superior temporal sulcus (STS) and bilateral temporo-parietal junction (TPJ) when subjects made choices facing human opponents rather than a computer. The foci of activity in the mPFC (peak MNI coordinates, x = 0, y = 48, z = 24) are consistent with results of many studies on theory of mind or mentalizing (see **Fig. 1**; Fletcher et al., 1995; Gallagher et al., 2000; McCabe, Houser, Ryan, Smith and Trouard, 2001; Bird, Castelli, Malik, Frith and Husain, 2004; Amodio and Frith, 2006). Psychologists and philosophers define as theory of mind or mentalizing, the ability to think about others' thoughts and mental states in order to predict their intentions and actions.

---Figure 1 about here---

When we analyzed separately high- and the low-level reasoning subjects, we found the activity in the medial prefrontal cortex to be stronger in subjects classified as high level (**Fig. 2**). In the high reasoners, guessing a number in the human condition activated two main regions of the medial prefrontal cortex, a more dorsal and a more ventral portion of the anterior mPFC.

---Figure 2 about here---

The prefrontal activity of the low-level reasoning subjects was found in the rostral anterior cingulate cortex (**Fig. 2**) (see section 3.4 below for an interpretation of the data).

fMRI results show additional brain activities related to high- versus low-level reasoning in the right and left lateral orbitofrontal cortex and left and right dorsolateral prefrontal cortex, areas likely related to performance monitoring and cognitive control (Koechlin and Summerfield, 2007). This suggests that a complex cognitive process subserves the higher level of reasoning about others.

The beauty contest game also requires solving a complex calculation task. Thus, in order to follow a first or higher level of reasoning, the subjects need to mentally multiply what they think might be the average of the numbers guessed by the others, 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, the inferior parietal lobule, and the supramarginal gyrus, was found both in the human and computer conditions. Results from our calculation task show enhanced activity in the angular gyrus and in the inferior parietal lobule when the subjects were requested to mentally multiply a factor times a number (C1 condition), and greater activity in the same areas 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. Notably, no activity of the medial prefrontal cortex was related with any kind of calculation.

3.3 The medial prefrontal cortex correlates with Strategic IQ

In Coricelli and Nagel (2009) we found a cross-subject correlation between a measure of strategic IQ in the beauty contest (computed as the distance of own choice to the target number, M*average of all chosen numbers, across all trials) and brain activity in the mPFC. 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. Strategic IQ was not correlated with accuracy (number of exact responses) in the calculation task, thus it is independent of cognitive or calculation skills. Notably, no other brain region of interest was correlated with strategic IQ. This suggests that the mPFC, involved in higher reasoning about others, leads to successful outcomes in our interactive setting.

3.4 Dorsal and ventral medial prefrontal activity: self-other distinction

As described above (**Fig. 2**), we found two portions of the medial prefrontal cortex, a more dorsal and a more ventral one, which are activated in human vs. computer condition for high level or reasoning only. The foci of activity in the medial prefrontal cortex are consistent with results of many studies on mentalizing (Fletcher et al., 1995; Bird, 2004; Gallagher et al., 2000; McCabe, et al., 2001). The underlying processing of high level of reasoning in the guessing game implies thinking about others thinking of you thinking about them, and so on; this implies that the higher level of reasoning subjects considered the others potentially ‘like them’. In other words they assume that the same reasoning that they are performing is likely performed by others, thus inducing a process of iterative thinking towards higher levels of resoning. This process implies that they deeply think about others in a ‘like me’ fashion. As shown in previous neuroimaging studies (Mitchell,

Banaji, Macrae, 2005; Mitchell, Macrae, Banaji, 2006), judging if others are similar to self activates the ventral anterior medial PFC. Moreover, third person perspective (put yourself in the shoes of the other) in making judgement about self mediates activity in the medial prefrontal cortex (D'Argenmbreu et al., 2007). Our results suggest that those two types of mental processing characterized higher level of reasoning in our experimental guessing game. Thus, deep strategic thinking implies both considering the others as like minded, and taking a third person perspective of our own behaviour. The main prefrontal activity of the low level of reasoning subjects was found in the rostral anterior cingulate cortex, an area often attributed to self-referential thinking in social cognitive tasks (Moran, Macrae, Heatherton, Wyland and Kelley, 2006). Thinking about the others as random players, thus considering them as 'zero-intelligent' agents needs only a first person perspective of the interactive context.

3.5 Pattern of neural activity related with recursive thinking

fMRI results show additional brain activities related to high versus low level of reasoning in the right and left lateral orbitofrontal cortex and bilateral BA44. The involvement of those areas suggests that higher level of reasoning requires the use of a complex cognitive apparatus. Lateral orbitofrontal cortex (BA 47) is often related with switching in cognitive states, which in our experimental task might refer to switching from thinking about self and others thinking about you, and so on. BA 44 might be related with the sequencing component of recursive thinking needed in higher level of reasoning.

5. Theory of mind (Mentalizing) and strategizing

We hypothesize that strategizing relies heavily on a Theory-of-Mind Mechanism (ToMM) or mentalizing. Thinking of other's mind is a normal ability of our species that has evolved over time as a result of social (interpersonal) interactions. Whether or not a Theory of Mind is an exclusive human ability is still an open question. (See: Baron-

Cohen ,1995; Woodruff and Premack, 1979; Gallup, 1970; Tomasello, Kruger and Ratner, 1993; Povinelli, 1993).

In Coricelli and Nagel (2009) we design an experiment to test whether or not a Theory of mind Mechanism is activated during strategic interactions and if this mechanism is related with depth of strategic reasoning. In the psychological and philosophical literature there are two main theories about theory of mind or mentalizing. The so-called "Theory-Theory" approach assumes that we use a simplified theory of human behavior when we attribute mental states or beliefs to others in order to predict their actions. According to the second approach, called simulation theory, people predict and interpret the behavior of others by imagining being in their situation (in terms of their mental state). Individuals "put themselves in the other's shoes," (Gordon, 1995). Simulation theory states that we predict and explain the behavior of other individuals by a simulative process, i.e. we simulate the decision-making process of the other individual by using part of our cognitive systems (Goldman, 1995; Gordon, 1995). "The simulation approach postulates that the heuristics or material employed in mentalizing make essential use of the attributer's own psychology. In the standard lore of simulation theory, an attributer who wishes to predict a target's decision begins by creating pretend states in himself that correspond (or so he thinks) to prior states of the target. He feels these pretend states into his own decision-making mechanism, and sees what decision the mechanism outputs." (cf. Goldman, 2001, p. 2). According to simulation theory we simulate the mental states of the other individuals using our own decision- making mechanism. This process is domain specific, considering that our decision-making mechanisms are different and specialized for different contexts. Degrees of knowledge of the others and the context (Coricelli, McCabe, and Smith, 2000), ranging from certainty to uncertainty; and the different levels of recursive reasoning (depths of reasoning), are crucial factors in the definition of the brain circuits that are needed to solve the interactive situation. This

approach is in contrast with the existence of a single theory-of-mind module, and calls for future studies aimed at understanding the underlying complexity of the mechanisms that drive social interaction.

In our experiment we could distinguish two behavioral types in terms of their levels of reasoning. Low level of reasoning subjects played in the same way with human or computer opponents, indeed they played level 1 in both conditions. They best respond to their beliefs that others (either humans or computer) would play randomly. Thus, low level of reasoning subjects used a simplified model of others' behavior (**Fig. 3**).

---Figure 3 about here---

In contrast, high level of reasoning subjects best responded to the beliefs that the others would play at level 1 (or higher). This implies that in defining their beliefs about others' behavior they used their own decision-making procedure (best response). They indeed assume that also other players best respond to their beliefs about other players' behavior (**Fig. 3**). This suggests that low level might have a simplified model of others' behaviour which can be interpreted that they have no model of other players' thinking process or they use a simplified statistical model as if they were playing against nature, while high level of reasoning might simulate the behavior of others with their own decision making procedure. This interpretation of the possible mental processes underlying the observed behavior in our experiment fits quite well the observed pattern of brain activity related to the low and high level of reasoning subjects (i.e., the activity of brain structure related to complex cognitive functions during high strategic reasoning).

6. Learning and strategic reasoning

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 own behaviour can influence others' behaviour, as reported in a recent study (**Fig. 4**) (Hampton et al., 2008).

---Figure 4 about here---

In the study by Hampton et al. the activity in the mPFC is found when contrasting two dynamic models of choice in a repeated competitive game. One based on updating own strategy based on other's past choices (Fictive), giving best response to the frequency play of actual behaviour, is essentially our level 1 thinking. A second, more sophisticated type, assumes that subjects considered the effect of their own past choices on other's behaviour (Influence). The contrast therefore is analogous to the difference in the beauty contest game between level 2 (or higher) and level 1 of strategic reasoning.

Thus, the mPFC encoding the effect of our choices on others' thought and behaviour is the neural signature of high level of strategic reasoning (level 2 or more). The main difference between these two studies are that in Hampton et al. subjects observed others' behaviour over time and need to respond to it, while in our study the subjects need to model also the choices of the others. The brain does not seem to distinguish between these two data sources. Taken together, the results of these two studies represent the first close link between adaptive learning and levels of reasoning.

7. How neuroscience can inform economics: specifications of the underlying processing of human's out-of-equilibrium behaviour

In the experimental beauty contest game, levels of reasoning were not induced (unlike the tasks used by (Bhatt and Camerer, 2005; D'Argembeau et al., 2007)), and we could detect heterogeneity between subjects based on their own choice of depth of reasoning. The main finding of the study by Coricelli and Nagel (2009) is that the mPFC clearly

distinguishes high- versus low-level of strategic reasoning, thus encoding the complexity underlying human interactive situations.

The pattern of brain activity in the right and left lateral orbitofrontal cortex and in the dorsolateral prefrontal cortex suggests a substantial jump in complexity when going from first to 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. This result provides a new interpretation that should be implemented in game theoretical modelling. This important difference has never been discussed in the experimental economics literature on strategic reasoning. Instead, the main difference has been thought to be between random behaviour and higher level; mainly because level 1 contains already best reply structure, a fundamental concept in economic theory. However data from Coricelli and Nagel (2009) show that the main discontinuity is in the belief about other's behaviour as naïve or random behaviour (the underlying belief of level 1 players) vs. belief of best reply behaviour (level 2 or higher).

Rational game theory only predicts equilibrium play, supposing common knowledge of rationality - everybody is rational and thinks that everybody else is rational, and so on. However actual behaviour deviates from equilibrium. In fact, humans use bounded rational strategies or cognitive hierarchies to mimic optimal behaviour. Thus, people behave differently based on different beliefs about others' behaviour. The results of our study demonstrate that much of the variation in strategic behaviour lies in individuals' different attitudes towards others. Crucially, behaviour that was based on more self-referential thinking ("I believe that others just play randomly") resulted in a larger deviation from rationality. Thus, people who are socially and strategically more intelligent are likely to reason in a less self-referential way i.e. they incorporate that others are also reasoning.

This paper should be seen as a contribution to McCabe's statement: "*Herbert Simon's research on bounded rationality (Simon, 1957) implies that strategies are likely to be encoded in the brain as a mapping from partitions of circumstances into partitions of actions together with inferential (*Holland, Holyoak, Nisbett, *1986) and reasoning mechanisms (Gigerenzer and Selten, 2001) that modify and scale these partitions. To understand how such encodings and mechanisms are formed requires both a top down approach using experimental methods [experimental beauty contest] and strategic models from economics [cognitive hierarchy model] and a bottom up approach using experimental methods [fMRI beauty contest] and computational models from cognitive neuroscience"*.

References

Amodio, D.M., Frith, C.D. (2006). Meeting of minds: the medial frontal cortex and social cognition. Nature Review Neuroscience, 7, 268-277.

Baron-Cohen, S. (1995). Mindblindness, An Essay on Autism and Theory of Mind. Brandford Books, MIT Press.

Bhatt, M., Camerer, C.F. (2005). Self-referential thinking and equilibrium as states of mind in games: fMRI evidence. Games and Economic Behavior, 52, 424-459.

Bird, C.M., 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.

Bosch-Domenech, A., Montalvo, J.G., Nagel, R., Satorra, A. (2002). One, Two, (Three), Infinity, ... : Newspaper and Lab Beauty-Contest Experiments. American Economic Review, 92,1687-1701.

Camerer, C.F., Lovallo, D. (1999). Overconfidence and Excess Entry: Experimental Evidence. American Economic Review, 89, 306-318.

Camerer, C.F., Ho, T-H., Chong, J-K. (2004). A Cognitive Hierarchy Model of Games. The Quarterly Journal of Economics, 119, 861-898.

Coricelli, G., McCabe, K., Smith, V. (2000). Theory-of-mind mechanism in personal exchange. In Hatano, et al. (Eds.), Affective minds (pp. 249–259). Elsevier Science.

Coricelli, G., Nagel, R. (2009). Neural correlates of depth of social reasoning in medial prefrontal cortex. Proceeding of the National Academy of Science USA, 106, 9163-9168.

Costa-Gomes, M., Crawford, V.P. (2006). Cognition and Behavior in Two-Person Guessing Games: An Experimental Study. American Economic Review, 96.

D'Argembeau, A., Ruby, P., Collette, F., Degueldre, C., Balteau, E., Luxen, A., Maquet, P., Salmon, E. (2007). Distinct regions of the medial prefrontal cortex are

associated with self-referential processing and perspective taking. Journal of Cognitive Neuroscience, 19, 935-944.
Fletcher, P.C., Happe, F., Frith, U., Baker, S.C., Dolan, R.J., Frackowiak, R.S., Frith, C.D. (1995). Other minds in the brain: a functional imaging study of "theory of mind" in story comprehension. Cognition, 57, 109-128.
Gallagher, H.L., Happe, F., Brunswick, N., Fletcher, P.C., Frith, U., Frith, C.D. (2000). Reading the mind in cartoons and stories: an fMRI study of 'theory of mind' in verbal and nonverbal tasks. Neuropsychologia, 38, 11-21.
Gallup, G. G., Jr (1970). Chimpanzees: Self-recognition. Science, 167, 86-87.
Gigerenzer, G., and Selten, R. (2001). Bounded rationality: The adaptive toolbox. MIT Press.
Goldman, A. (1995). Interpretation Psychologized ? In M. Davies & T. Stone (Eds.), Folk psychology: The theory of mind debate. Oxford: Blackwell.
Goldman, A. (2001). Using your mind to read others. University of Ari- zona Working Paper.
Gordon, R. M. (1995). Folk psychology as simulation? In M. Davies, T. Stone, & Folk (Eds.), Psychology: The theory of mind debate. Oxford: Blackwell.
Hampton, A.N., Bossaerts, P., O'Doherty, J.P. (2008). Neural correlates of mentalizing-related computations during strategic interactions in humans. Proceeding of the National Academy of Science USA, 105, 6741-6746.
Ho, T-H., Camerer, C.F, Weigelt, K. (1998). Iterated dominance and iterated best response in experimental "p-Beauty contests". American Economic Review, 88, 947-969.
Holland, J.H., Holyoak, K.J., Nisbett, R.E. (1986). Induction: processes of inference, learning, and discovery. MIT Press.
Keynes, J.M. (1936). The General Theory of Employment Interest and Money. Cambridge University Press: Macmillan.
Koechlin, E., Summerfield, C. (2007). An information theoretical approach to prefrontal executive function. Trends in Cognitive Science, 11, 229-235.
McCabe, K., Houser, D., Ryan, L., Smith, V., Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. Proceeding of the National Academy of Science USA, 98, 11832-11835.
Mitchell, J.P., Banaji, M.R., Macrae, C.N. (2005). The link between social cognition and self-referential thought in the medial prefrontal cortex. Journal of Cognitive Neuroscience, 17, 1306– 1315.
Mitchell, J.P., Macrae, C.N., Banaji, M.R. (2006). Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. Neuron 50, 655– 663.
Moran, J.M., Macrae, C.N., Heatherton, T.F., Wyland, C.L., Kelley, W.M. (2006). Neuroanatomical evidence for distinct cognitive and affective components of self. Journal of Cognitive Neuroscience, 18, 1586-1594.
Nagel, R. (1995). Unreveling in Guessing Games: An experimental Study. The American Economic Review, 85, 1313-1326.
Povinelli, D. J. (1993). Recostructing the evolution of mind. American Psychologist, 48, 493-509.
Selten, R. (1998). Features of experimentally observed bounded rationality. European Economic Review, 42, 413-436.

Simon, H.A. (1957). Models of Man: Social and Rational. New York: Wiley.
Stahl, D.O., Wilson, P.W. (1995). On Players'Models of Other Players: Theory and Experimental Evidence. Games and Economic Behavior, 10, 218-254.
Tomasello, M., Kruger, A., and Ratner, H.H. (1993). Cultural learning. Behavioral and Brain Sciences, 16, 495-552.
Woodruff, G., and Premack, D. (1979). Intentional communication in the chimpanzee: the development of deception. Cognition, 7, 333-362.

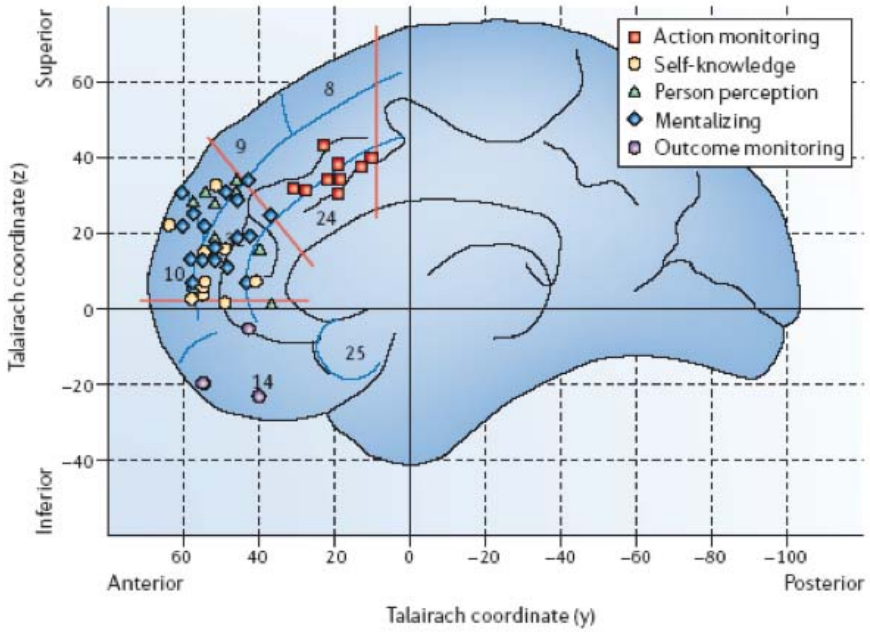


Fig 1. Each dot in this template (human brain) represents a focus of activity related to different functions (e.g. action monitoring, self-knoledge, etc.) found in independent neuroimaging studies. Activity related to mentalizing is found in the medial prefrontal cortex (Brodmann areas BA10 and BA32, also called paracinglulate). This activity is found when contrasting mentalizing vs. non mentalizing tasks, thus when the participants are asked to ascribe and attribute mental states and beliefs to others to interpret and understand their behavior vs. tasks in which the understanding of the context does not require any attribution of mental states. When the task is an experimental game, the contrast often used to isolate mentalizing activity is human-human vs. human-computer interaction (Coricelli, McCabe and Smith 2000; McCabe et al 2001). The figure is reproduced from Amodio and Frith (2006), with permission.

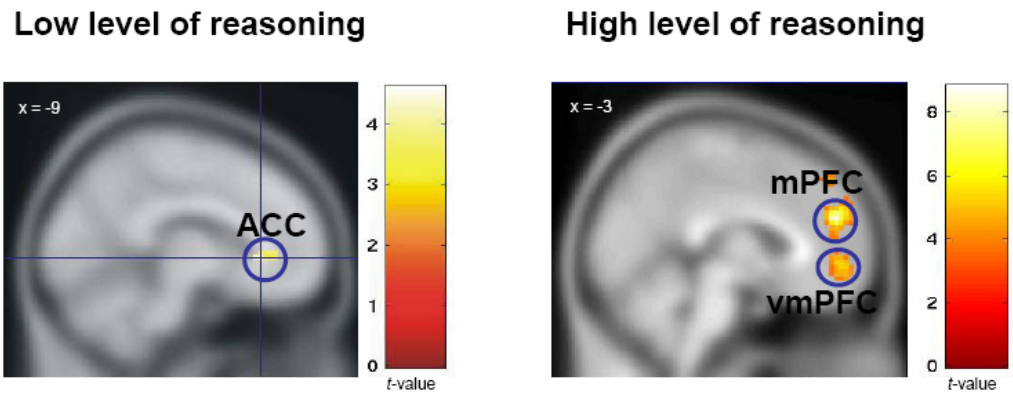


Fig. 2 Pattern of neural activity related to low and high level of reasoning in the Beauty Contest game. Guessing in the human condition in contrast to the computer condition was associated with relative enhanced activity in the rostral anterior cingulate cortex (**Left panel**, low level of reasoning subjects, ACC); and (**Right panel**, high level of reasoning subjects) activity in the dorsal medial prefrontal cortex (mPFC) and ventral medial prefrontal cortex (vmPFC). This shows how the neural activity of low and high level of reasoning people differs.

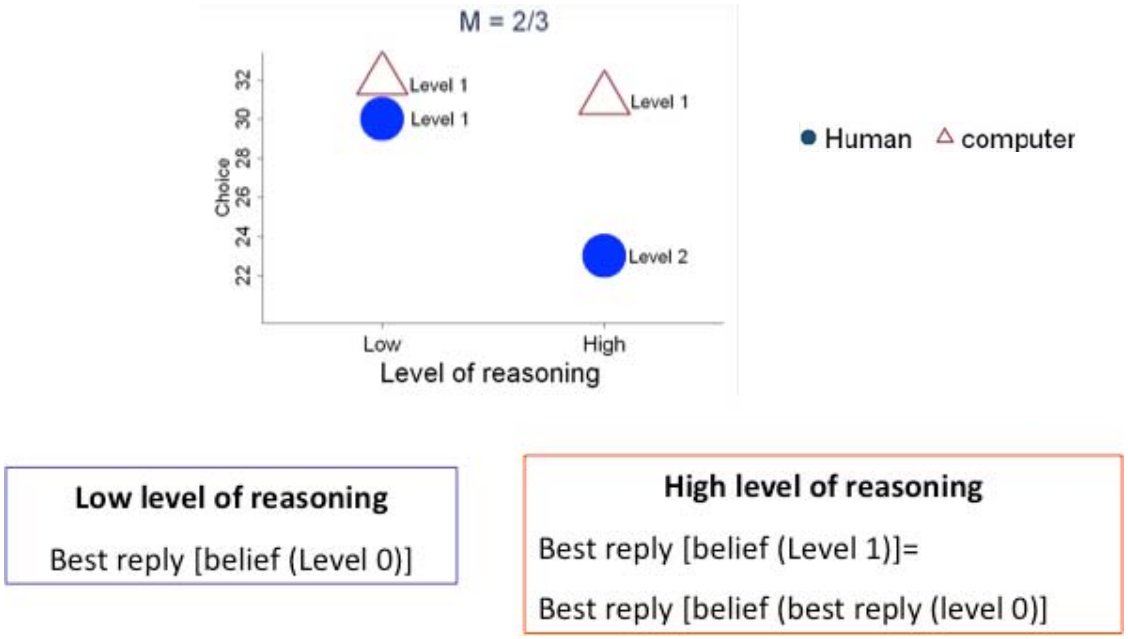
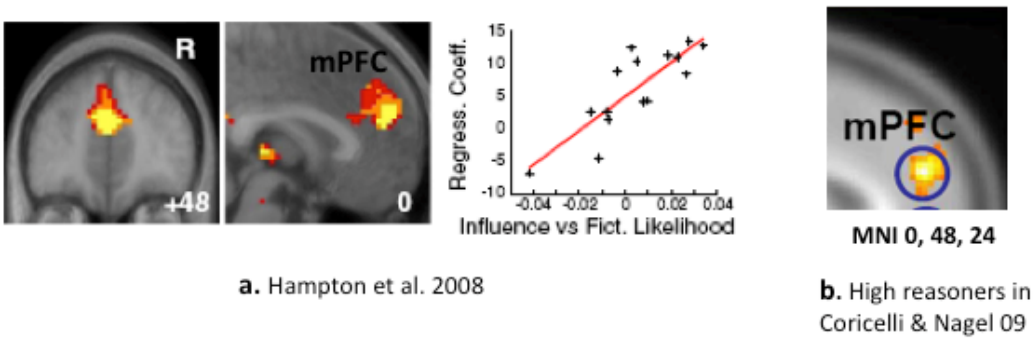


Fig 3. The subjects classified as low level behaved similarly against the computer or the humans, thus they behaved as level 1 in both conditions (**Left**). The high level of reasoning subjects clearly differentiated their behaviour in the human compared to the computer condition (**Right**). They behaved as level 1 in the computer condition while being classified as higher level of reasoning (level 2 or more) when interacting with human counterparts. Low level of reasoning implies a best reply to the belief that others will play Level 0 (i.e. will play randomly); while, high level of reasoning implies a best reply to the belief that others will play at Level 1 (or higher), this means that high level of reasoning subjects will use their own decision making procedure (i.e. best reply) to compute the beliefs about the behaviour of the others.



a. Hampton et al. 2008

b. High reasoners in Coricelli & Nagel 09

Fig 4. a. In the study by Hampton et al. the activity in the mPFC is found when contrasting two dynamic models of choice in a repeated competitive game. One based on updating own strategy based on other's past choices, giving best response to the frequency play of actual behaviour (Fictive). A second, more sophisticated type, assumes that subjects considered the effect of their own past choices on other's behaviour (Influence). Activity (betas) in the mPFC correlates with the (across subjects) difference of the likelihoods between Influence and Fictive. **b**. The same mPFC activity is found in Coricelli and Nagel 2009 for high reasoners in the beauty contest game. Adapted from Hampton et al. (2006).