NeuroImage 53 (2010) 790–799

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

The neural mechanisms of learning from competitors

Paul A. Howard-Jones [a,*], Rafal Bogacz [b], Jee H. Yoo [b], Ute Leonards [c], Skevi Demetriou [a]

[a] Graduate School of Education, University of Bristol, 35 Berkeley Square, Bristol, BS8 1JA, UK
[b] Department of Computer Science, University of Bristol, Merchant Venturers Building, Woodland Rd, Bristol, BS8 1UB, UK
[c] Department of Experimental Psychology, University of Bristol, 12a, Priory Road, Bristol, BS8 1TU, UK

ARTICLE INFO

Article history:
Received 13 January 2010
Revised 24 May 2010
Accepted 9 June 2010
Available online 16 June 2010

ABSTRACT

Learning from competitors poses a challenge for existing theories of reward-based learning, which assume that rewarded actions are more likely to be executed in the future. Such a learning mechanism would disadvantage a player in a competitive situation because, since the competitor's loss is the player's gain, reward might become associated with an action the player should themselves avoid. Using fMRI, we investigated the neural activity of humans competing with a computer in a foraging task. We observed neural activity that represented the variables required for learning from competitors: the actions of the competitor (in the player's motor and premotor cortex) and the reward prediction error arising from the competitor's feedback. In particular, regions positively correlated with the unexpected loss of the competitor (which was beneficial to the player) included the striatum and those regions previously implicated in response inhibition. Our results suggest that learning in such contexts may involve the competitor's unexpected losses activating regions of the player's brain that subserve response inhibition, as the player learns to avoid the actions that produced them.

Introduction

Learning from competitors is a critically important form of learning for animals and humans. Animals must frequently compete for mates, resources and social dominance, and outcomes can be critical for their reproduction or even survival. Young animals often engage in play fighting that is regarded as preparatory for their struggles in adult life (Groos, 1898). To fully benefit from such play fighting, young animals need to learn from their competitors' failures and successes.

This type of learning presents a challenge for existing neuropsychological concepts and for the computational models associated with them. Current understanding of animal and human reinforcement learning rests on the assumption that rewarded actions are more likely to be executed in the future. This simple association between reward and action appears less helpful in understanding the behavior of a participant in a competitive situation (referred to here as the *player*). From the player's perspective, an action by the competitor that results in the competitor's loss provides a reward (de Bruijn et al., 2009). However, this is counter to how the player should value the action producing it, when making their own decisions in the future. Standard reinforcement learning models thus appear inadequate in competitive contexts, since they suggest a competitor's unexpected losses would produce reward activity favouring actions better avoided. Similarly, a competitor's unexpected gains would not provide the types of immediate reward likely to encourage the player to repeat the actions that produced them.

The challenge posed by learning from competitors becomes clearer when we consider the neuro-computational models of reinforcement learning. The models assume that the player maintains estimates of values of individual actions, which we denote by m_i. After selecting action i, the corresponding action value is modified proportionally to the player's prediction error (Dayan and Abbott, 2001).

$$m_i \leftarrow m_i + \eta\delta_p, \text{ where } \delta_p = r_p - m_i \quad (1)$$

In Eq. (1), η is a learning rate, and player's prediction error δ_p is the difference between reward r_p obtained by the player and the expected reward m_i. Thus for example, if the reward obtained is larger than predicted, then $\delta_p > 0$ and, according to Eq. (1), the value of the chosen action is increased. It has been proposed that δ_p is represented in the firing rate of dopaminergic neurons (Montague et al., 1996) and strong experimental support has been provided for this theory (D'Ardenne et al., 2008; Schultz et al., 1997; Tobler et al., 2005; Zaghloul et al., 2009). It has been further proposed that m_i are encoded in strengths of synaptic connections which are modified according to the level of dopamine as in Eq. (1) (Montague et al., 1996). This is supported by observation of the effects of dopamine on synaptic plasticity in the striatum (Reynolds et al., 2001; Reynolds and Wickens, 2002), where midbrain dopaminergic neurons predominantly converge. Such a proposed mechanism is also supported by observations that the activity in reward-related regions measured

* Corresponding author. Fax: +44 117 925 1537.
E-mail addresses: Paul.howard-jones@bristol.ac.uk (P.A. Howard-Jones), R.Bogacz@bristol.ac.uk (R. Bogacz).

1053-8119/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2010.06.027

with fMRI is correlated with δ_p (Daw et al., 2006; McClure et al., 2003; O'Doherty et al., 2004).

Developing a hypothesis for how a player learns from their competitor's feedback is less straightforward. This may occur by updating estimates of action values in a manner analogous to that for their own feedback. That is, when a competitor selects action i and receives reward r_c, the player may update their estimate of the action value as follows:

$$m_i \leftarrow m_i + \eta\delta_c, \text{ where } \delta_c = r_c - m_i \quad (2)$$

We refer to δ_c as the competitor's prediction error. Note that the competitor's prediction error is low when the competitor chooses an unsuccessful action, but such outcome is beneficial for the player. Hence, it may seem unlikely that δ_c is encoded in the firing rate of dopaminergic neurons. A recent neuroimaging study (de Bruijn et al., 2009) suggests instead that dopaminergic neurons may exhibit an opposite pattern of response and show high activity for the competitor's losses. Hence, we define egocentric prediction error as $\delta_e = -\delta_c$. Note, however, that if the egocentric prediction error is encoded in the firing rate of dopaminergic neurons, then action values m_i should *not* be modified proportionally to δ_e, because this would increase values of actions that gave the competitor low rewards, as mentioned at the start of this paper.

To understand the mechanisms by which the human brain learns from competitors, we performed an experiment in which players competed with a computer in a task in which they could gain points by selecting one of four bandits on each trial (Fig. 1). We tested the following hypotheses. First, we compared the ability of players to learn from their competitors' outcomes relative to the outcomes of their own actions. Despite the challenge that learning from competitors presents for current models, we hypothesized that the players would be able to use the information provided by the competitor's feedback to guide their own choices, because this ability provides advantage in competitive situations.

Second, we investigated how the brain represents variables required for learning from competitors, in particular, those representing the competitor's actions. Since our task involved making different choices with different hands, we compared the activities in hand-specific regions when the player performed an action with those generated when observing the competitor. Among regions included in the mirror neuron system (MNS), premotor regions and primary motor cortex are likely to be hand-specific. There is considerable evidence for primary motor cortex activation in response to observed action (Caetano et al., 2007; Cochin et al., 1999; Hari et al., 1998). For example, it responds to observed hand movement in the hemisphere contralateral to the hand image provided as stimulus (Touzalin-Chretien and Dufour, 2008). On this basis, we hypothesised that when players observed their competitor's action, they would activate premotor and motor cortex in the same regions activated as when making the action themselves, i.e. contralateral to the handedness of the action.

Third, on the basis of previous studies of reinforcement learning (reviewed above), we hypothesized that, when the player is learning from their own outcomes, ventral striatal activity would be positively correlated with the players' own prediction error (δ_p).

Fourth, we tested two opposing hypotheses: that the prediction error (either the competitor's prediction error (δ_c) or the egocentric prediction error (δ_e)) would be encoded in the ventral striatum when observing the competitor.

Fifth, on the basis that our results would support the latter hypothesis (as was the case) we predicted activity correlated with δ_e in regions previously linked to response inhibition, since the potential reward provided by the competitor's loss can only be reaped by suppressing the action that gave rise to it. The key regions of interest were those most consistently identified in the literature: the right

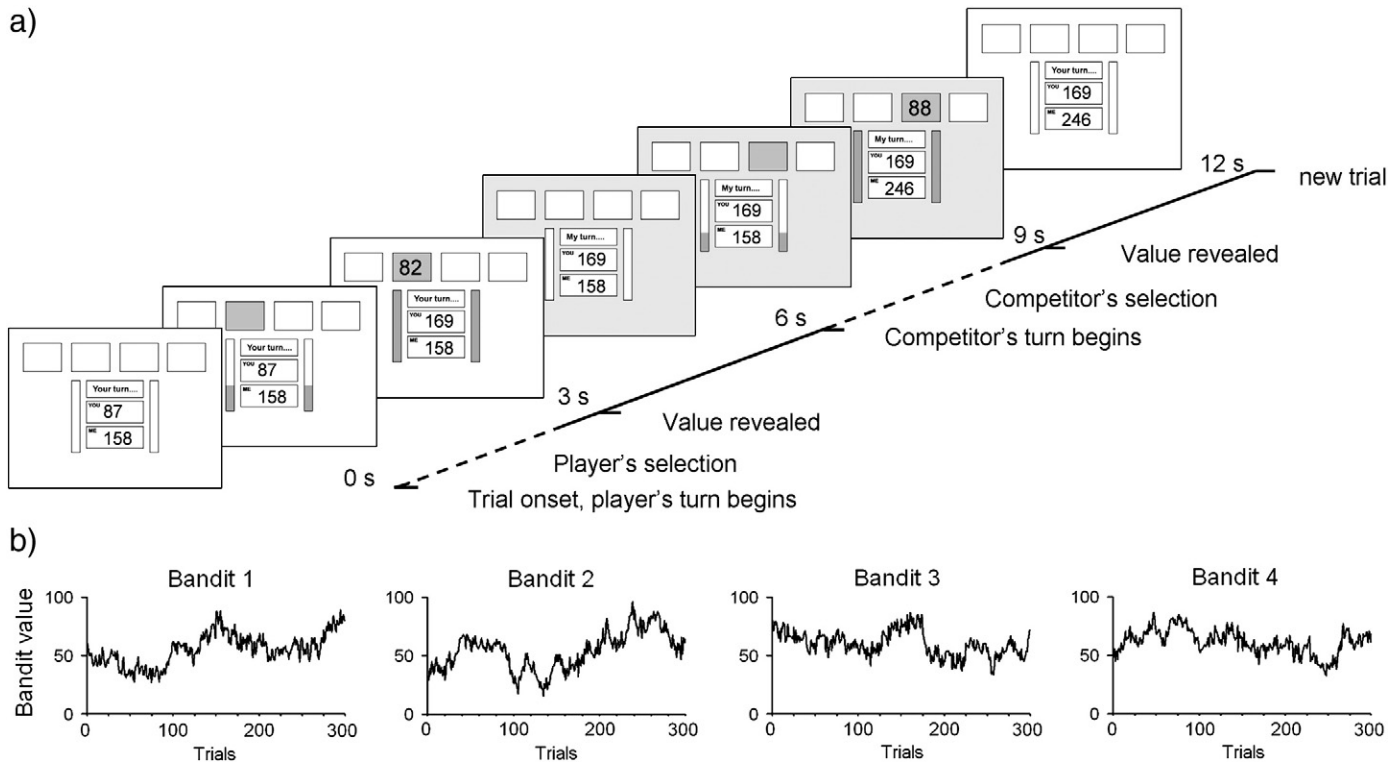

Fig. 1. Task design a) Illustration of the timeline within a trial (see Materials and methods for description of trial). b) Example of payoffs that would be received by choosing each bandit box on each trial.

inferior frontal gyrus (IFGr) (Aron et al., 2007; Aron and Poldrack, 2005, 2006; Cai and Leung, 2009; Coxon et al., 2009; Garavan et al., 2006; Garavan et al., 1999; Konishi et al., 1998; Leung and Cai, 2007; Pliszka et al., 2006; Rubia et al., 2001; Rubia et al., 2003) and right middle frontal gyrus (MFGr) (Aron et al., 2007; Cai and Leung, 2009; Garavan et al., 2006; 1999; Pliszka et al., 2006; Rubia et al., 2001).

Our sixth hypothesis, also assuming that the player would be monitoring the competitor's losses, predicted activity related to the value of the competitor's losses would be found in the posterior medial frontal cortex (pMFC), a region associated with detecting errors irrespective of whether these are self-made errors or made by a competitor(de Bruijn et al., 2009).

Finally, we investigated the neural correlates of the player's exploratory behavior i.e. decisions not to exploit the bandit with the greatest m_i. With regard to activity correlated with exploratory decision-making, we hypothesised activity in prefrontal cortex as the region most implicated in the type of behavioral control required for switching strategies (Daw et al., 2006; Miller and Cohen, 2001).

Our findings suggest that the player represents the competitor's actions in a similar way to their own, with the competitor's unexpected losses engaging the player's reward and response inhibition systems, as they learn to avoid the actions that produced them.

Materials and methods

Participants

Players were 16 healthy participants between 20 and 34 years old (8 male and 8 female, mean age 25.5/SD 3.8 years) who provided written informed consent and were right handed as assessed by the Edinburgh Handedness Inventory.

Task

The task was an adapted form of the 4-armed bandit task (Daw et al., 2006), in which players alternated turns with a computer competitor in selecting one of four bandits (see Fig. 1a). Four boxes, representing the bandits, were displayed on the screen and players were given 3 s to make a decision, with the time elapsed shown on two bar indicators that were positioned symmetrically to the left and right of centre. The player always began the game. He/she was asked to indicate their choice of bandits by pressing one of two buttons held in the left hand (for bandits 1 and 2) or right hand (for bandits 3 and 4). As soon as a bandit was selected, it would change color from purple to red to indicate a decision had been made. At the end of the decision window, the pay-out was displayed for another 3 s in the centre of the bandit. This value then disappeared, the selected bandit returned to purple and the background became slightly darker, to indicate that the competitor was now playing. The computer competitor made a selection within the next 3 s, at a randomly selected instant between 500 ms and 2000 ms from the beginning of the competitor's decision window. As with the player's turn, the bandit selected by the competitor changed color immediately, and the outcome of the competitor's selection was revealed as soon as the decision window had elapsed and this outcome was displayed for a further 3 s. The next 12 s trial then began immediately. Boxes in the centre of the screen displayed total scores and bars on either side indicated the fraction of decision window elapsed.

The payouts from *i*th bandit were generated as in the study of Daw et al. (2006). They were positive integers between 1 and 100. They were drawn from a Gaussian distribution with mean μ_i and standard deviation $\sigma_o = 4$, and rounded to the nearest integer. After each trial, the means for each *i* diffused in a decaying Gaussian random walk:

$$\mu_i \leftarrow \lambda\mu_i + (1-\lambda)\theta + v. \quad (3)$$

In Eq. (3), decay parameter λ was set at 0.9836, decay centre θ set at 50 and diffusion noise v was sampled from a Gaussian distribution with mean zero and standard deviation $\sigma_d = 2.8$ (see Fig. 1b for an example of μ_i used).

The competitor's behavior followed an *e*-greedy model (Sutton and Barto, 1998). According to this model, the competitor kept track of the most recently obtained reward for each bandit. On each trial the competitor selected, with probability $(1-e)$, the bandit whose most recent reward was highest or, with probability e, made a random selection (from a uniform distribution). The competitor's exploit/explore behavior was set to provide a suitably challenging level of difficulty for the participating players to compete with. Preliminary trials helped to inform this decision, and these were carried out in a simulated scanner to help acclimatize players to the scanning environment. In these trials, all players experienced the task with the competitor's e set at 3 levels of 0.5, 0.25 and zero probability of making a random selection, permutated with 3 instantiations of the above process for generating bandit values. Players experienced 75 trials in each condition, with all 16 players winning in the $e = 0.5$ condition, 9 players winning in the $e = 0.25$ condition and only 1 player succeeding in overcoming their competitor when e was set at zero. The $e = 0.25$ setting was, therefore, selected for the competitor in the imaging study, since the performance of the competitor at this setting best matched that of the players.

Procedure

Players were unpaid volunteers but, to ensure they remained competitively engaged with the competitor, an inducement of £50 cash was offered to the player who, on any one day of scanning, beat the competitor by the greatest margin. (The scanning took place over three days, and so three of these awards were made.) Before entering the scanner, the players were shown the money they could win.

In the scanner, each player experienced two consecutive 30 min sessions, each consisting of 150 trials, in which they attempted to outperform their competitor and maximise their lead. Two instantiations of mean bandit payoffs μ_i (obtained from Eq. (3)) were used to generate bandit values (as in Daw et al., 2006), and their presentation order was balanced within and between the two subgroups formed by gender.

Imaging procedure and pre-processing of image data

Imaging was performed with a 1.5 T whole-body magnetic-resonance imager (Phillips Gyroscan Intera with quadrature head coil). The head of the player was strapped firmly but comfortably in the head coil. Attached to the head coil was a mirror through which could be viewed the projection of a computer screen, positioned beyond the bore of the imager. A T2* sensitive (BOLD) echo-planar imaging sequence was used for functional imaging with TR = 3000 ms, TE = 50 ms. For each of the two 30 min sessions experienced by each player, 600 3D volume acquisitions were obtained. Each 3D volume acquisition consisted of 32 contiguous slices, 64×64 matrix, with a voxel size of $3 \times 3 \times 3$ mm^3, in an oblique axial plane that was rotated 20° with respect to the anterior commisure–posterior commisure line to enable whole-brain coverage. The beginning of each 12 s trial was automatically initiated by the scanner.

Processing of data was performed off-line using SPM5 (Wellcome Department of Cognitive Neurology, UCL, London) using the computational facilities of the Advanced Computing Research Centre, Bristol University. Data from each player was first realigned to the first scan and "unwarped" using a model for susceptibility-by-movement interactions to remove the residual movement related variance (Andersson et al., 2001). The data for each player was then spatially normalised to the Montreal Neurological Institute template. Images were smoothed with a Gaussian kernel filter using a relatively large

12 mm FWHM, reflecting our intention to make group inferences with 16 subjects (Mikl et al., 2008). A temporal high-pass filter (128 s) was applied to remove low-frequency extraneous effects, such as cardiac and respiratory artefacts.

Computational models

In this study of competitive learning, we considered the following 5 variants of a neuro-computational model for estimating values of prediction errors on individual trials.

The first variant is given by Eqs. (1) and (2) and additionally assumes that at the start of a block of trials all action values are equal to m_0. Thus, this variant has two free parameters m_0 and learning rate η. The free parameters of all variants of the model are summarized in Table 1. The second variant is similar to the first, but additionally assumes that all action values decay with time, i.e. at the end of each trial all action values are updated according to:

$$m_i \leftarrow \lambda m_i + (1-\lambda)\theta \tag{4}$$

In Eq. (4), λ denotes the rate of decay (the lower λ the faster decay), and θ denotes the value to which the action values converge (if the corresponding action is not chosen). The third variant is the Kalman filter model used by Daw et al. (2006) to describe learning in a similar task. Essentially, it assumes that the learning rate η is not constant within a block, but varies depending on the participant's estimates of how noisy and variable the rewards are. Since this variant did not provide the best fit to experimental data we do not present it in detail here (detailed description in Daw et al., 2006).

The fourth variant is similar to the second, but additionally assumes that the learning rates for player and competitor feedback in Eqs. (1) and (2) have different values η_p and η_c. Finally, the fifth variant is similar to the first, but assumes different learning rates for player and competitor feedback.

For all the variants we assume that probability of choosing action i depends on action values according to the softmax rule (that provided best fit to behavioral data from a similar experiment by Daw et al., 2006):

$$P(i) = \frac{\exp(\beta m_i)}{\sum_{j=1}^{4} \exp\left(\beta m_j\right)}. \tag{5}$$

In Eq. (5), β is a parameter controlling how deterministic the choice is: If $\beta = 0$, choices are made randomly, and the higher β, the higher the probability of choosing the action with maximum m_i. β is an additional free parameter for all of the variants.

Fitting of behavioral data

We now describe how each of the variants was compared with behavioral data. For each variant we were seeking its parameter values that provided best fit to the behavioral data. For each set of parameters we calculated m_i on all trials on the basis of responses and feedback in the experiment. An example of such a calculation is given in Supplementary Table 1. Using these m_i, we calculated the probabilities of selection of different action $P(i)$ predicted by the model for all trials. Then we calculated the likelihood of participants' choices given the model: If we denote the choice made on trial t by c_t, then the likelihood of this choice given the model is equal to $P(c_t)$ computed from Eq. (5) using m_i estimated for this trial (see Supplementary Table 1). The likelihood of data from T trials is given by:

$$L = \prod_{t=1}^{T} P(c_t) \tag{6}$$

Note that the likelihood L can be computed for any combination of values of the model's parameters. Hence L is a function of the model's parameters, and thus finding the parameters' values that provide best fit to experimental data is equivalent to finding the maximum of function L. We used the simplex optimization algorithm (Nelder and Mead, 1965) to find values of parameters maximizing the likelihood of Eq. (6). Initially, we tried to fit a separate model to each player (i.e. only trials from a single player were included in the product of Eq. (6)). However, this led to unreasonable estimates for the parameters (e.g. learning rates significantly above 1), which suggested that the amount of behavioral data from a single player was not sufficient to constrain the model. Hence, we followed the approach of Daw et al. (2006) and fitted a single model to data from all players (i.e. trials from all players were included in Eq. (6)) and set all parameters constant across players except for β which was a free parameter for each player.

Since different variants of the model have different numbers of parameters, we compared the variants by computing the Akaike Information Criterion (AIC) (Akaike, 1974) which "penalizes" the variants with a large number of parameters k.

$$AIC = -2\log L + 2k \tag{7}$$

The values of AIC for all variants considered are listed in Table 1. According to the Akaike criterion, variant 4 provided the best account for the behavioral data. The parameters of variant 4 which provided the best fit to behavioral data were: $m_0 = 53.3$, $\lambda = 0.89$, $\theta = 54.1$, $\eta_p = 0.87$, $\eta_c = 0.72$.

Analysis of image data

In the statistical analysis, each trial was modelled with 4 time points. The first two were the time of the selection made by the player (arbitrarily set to be midway between beginning of the trial and the response of the player signalling their selection — on average 208 ms from the beginning of the trial) and the time of presentation of the outcome (3 s after trial onset). The second two time points were when the player observed their competitor's selection (as signalled by one of the four bandits changing color) and the time of presentation of the outcome of their competitor's decision.

The choice of regressors was motivated by the hypotheses stated in the Introduction and these are listed in Table 2. Two analogous sets of regressors were generated for the player's and competitor's turns (regressors 1–5 and 6–10 respectively). The regressor for handedness of decision characterised each player's selection in terms of the hand used to make it, coded as 1 for *left* choices if it required the use of the left hand (bandits 1 and 2) or −1 for *right* choices if it required the right hand (bandits 3 and 4). In the case of the competitor's selection, this was the hand that would have been used by the player if he/she had made it. The choice type regressor classified decisions as either exploitative when the bandit with maximum m_i was chosen or as exploratory if an alternative was chosen. Choice probability was included in the regressors because it can reflect choice confidence,

Table 1
Free parameters and Akaike Information Criteria (AIC) for the five variants of the computational model considered. The additional parameters of variant 3 (Kalman filter) denote: σ_0 — initial estimate of standard deviation of action values, $\hat{\sigma}_d$ and $\hat{\sigma}_o$ — estimates of σ_d and σ_o defined in Materials and methods section (under and above Eq. (3)).

Variant	Free parameters	AIC
1	m_0, η	4884
2	$m_0, \eta, \lambda, \theta$	4358
3	$m_0, \lambda, \theta, \sigma_0, \hat{\sigma}_d, \hat{\sigma}_o$	4392
4	$m_0, \lambda, \theta, \eta_p, \eta_c$	4353
5	m_0, η_p, η_c	4883

Table 2
Time points and values of regressors.

	Regressor	Time point	Regressor's values
1	Player's choice	Midpoint between start of trial and choice of the player	1 for left choices, −1 for right choices
2	Choice type		1 for exploitative, −1 for exploratory
3	Choice probability		$P(c_t)$
4	Player's reward	Player's outcome revealed	r_p
5	Prediction error		δ_p
6	Competitor's choice	Choice of the competitor revealed	1 for left choices, −1 for right choices
7	Choice type		1 for exploitative, −1 for exploratory
8	Choice probability		$P(c_t)$
9	Competitor's reward	Competitor's outcome revealed	r_c
10	Prediction error		δ_c (which is equal to $-\delta_e$)

which may have been lower on exploratory trials. Inclusion of choice probability reduced the possibility that confidence-related changes in activation were reported for exploratory trials.

We used statistical parametric mapping to identify brain regions where activity was significantly correlated with these regressors. Regressors were convolved with the canonical hemodynamic response function and entered into a regression analysis against each player's fMRI data using SPM5. In this way, for each player, contrast images were created for activity correlated with each regressor. For inference at group level, these contrasts were subjected to a second-level analysis in which random effects group statistics were generated. T1 anatomical images were co-registered to mean functional EPI images for each player and normalised using EPI image normalisation parameters. The t-maps from the functional analysis were overlaid onto an average of the normalised structural image across players, with localisation of activity carried out with reference to an anatomical atlas (Duvernoy, 1999). Unless otherwise stated, we report activation in hypothesised regions that survived a height threshold of $p<0.001$ uncorrected with an extent threshold of 10 voxels (Forman et al., 1995). For the sake of completeness, activities beyond hypothesised regions are reported in figures and supplementary tables according to the same criteria. Where unhypothesised activities warrant consideration in the text, they are accompanied by a more conservative test ($p<0.0001$ and 10 voxel extent threshold) to guard against Type 1 errors. To test hypotheses regarding striatal activity in relation to prediction error (δ_p, δ_c, δ_e), a region of interest was defined within the nucleus accumbens (NAcc), on the basis of recent anatomical and functional studies (Neto et al., 2008) and confirmed by our anatomical images.

To investigate whether brain regions for the player's own actions and those of the competitor were detectable in conjunction, a separate factorial model at the second level was carried out. This model included regression fits for handedness of decision for both player and competitor. Conjunction null analyses (Friston et al., 2005) of regression fits for left versus right decisions, and right versus left decisions, were carried out with players' and competitor's data in conjunction. A region of interest (ROI) analysis was carried out on these two conjunction analyses, centred on the peak voxel in the player's motor cortex when observing the competitor's actions.

Results

Behavioral results

To compare the ability of players to learn from their competitors' feedback relative to their own, we estimated parameters of the reinforcement learning model (Eqs. (1) and (2)) from behavioral data. As described in detail above, we tested 5 variants of the model that differed in respect of which parameters were allowed to vary (between trials or conditions), and which were held constant. The model that provided the best fit to the behavioral data was variant 4, which allowed the learning rates η for player and competitor feedback to differ. These rates were calculated as equal to $\eta=0.87$ for the player's (Eq. (1)) and $\eta=0.72$ for competitor's feedback (Eq. (2)). The similarity between these two values implies that players learnt only slightly less from the competitor's feedback than from their own.

Neural representation of the player's and competitor's actions

The regions with higher activity when the player was making left versus right choices included hypothesised regions involved with the control of left hand movement: A cluster of increased activity was observed extending from right somatosensory regions into primary motor cortex and another cluster was observed in left cerebellum. Complimentary activations in the opposite hemispheres were observed when the player was making right, compared with left choices (see Supplementary Table 2 and Supplementary Fig. 1a).

We next sought activities in premotor and motor regions when players observed their competitor's selections. Regions with higher activity when the player observed the competitor making left choices versus right choices included the right primary motor cortex and dorsal premotor regions. Analogously, for observing the competitor's right versus left choices, activity was identified in left primary motor cortex (Supplementary Table 3 and Supplementary Fig. 1b). Thus, as we hypothesized, players' encoding of their own and the competitor's actions generated a similar pattern of activity. To confirm this similarity we performed a conjunction analysis (see Materials and methods). An ROI was defined by a 6 mm sphere centred on the peak voxel activated in the players' motor cortex (precentral gyrus in left and right hemispheres, −38 −22 62 and 38 −15 45 respectively) when observing their competitor's actions. ROI analysis confirmed statistically significant conjoined activity in this region for both the players' own actions and for observing those of their competitor, when these actions were both left ($p_{FWE\text{-}corr}=0.005$) and right handed ($p_{FWE\text{-}corr}=0.026$). Fig. 2 shows the region of conjoined activation ($p<0.001$ uncorrected) for left-handed decisions (which also shows activity at $p<0.005$ uncorrected, in order to show extent of activation).

Activity related to player's feedback

No activity was significantly positively or negatively correlated ($p<0.001$ uncorrected) with the magnitude of the points won by the player. In the dorsal striatum, activity correlated with δ_p was observed ($p<0.001$ uncorrected) in caudate regions (Supplementary Table 4). To determine whether ventral striatal activity was correlated with δ_p (as hypothesized in the Introduction on the basis of previous studies),

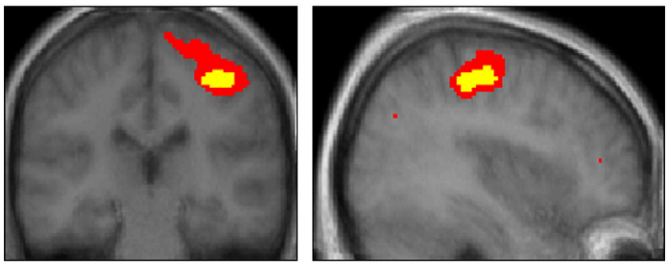

Fig. 2. Regions of overlap for representation of player's and competitor's left-handed actions, identified by conjunction null analysis of players' activation when they made left versus right decisions and when observing left versus right selections made by their competitor. Conjoined activation is shown that survives $p<0.001$ uncorrected (yellow) and, to illustrate full extent of activation, also $p<0.005$ uncorrected (red).

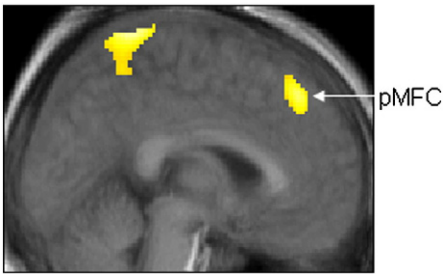

Fig. 3. Region of the posterior medial frontal cortex (pMFC, centre at x = −2, y = 39, z = 35) negatively correlated with the points won by the competitor (p<0.001 uncorrected).

we performed an ROI analysis on the NAcc. We defined two spheres of 4 mm radius at stereotactic coordinates (+/−10, 5.5, −4) derived from Neto et al. (2008), and confirmed by our anatomical images. This analysis revealed NAcc activity was correlated with δ_p in right ($p_{FWE\text{-}corr} = 0.001$) and left ($p_{FWE\text{-}corr} = 0.001$) hemispheres.

Activity related to competitor's feedback

Positive correlation of activity with the points won by the competitor did not reach threshold in any region of the brain but, as hypothesised, negatively correlated activity was revealed in the pMFC (Fig. 3, and Supplementary Table 5).

We wished to determine whether the activity indicative of dopaminergic response encoded δ_c or δ_e when players observed their competitor's feedback. To achieve this, we sought positive correlation between striatal activity and these two prediction errors. There were no brain regions where activity was positively correlated with δ_c (p<0.001 uncorrected). The same ROI analysis as described above revealed activity correlated with δ_e for the competitor's outcomes reached significance ($p_{FWE\text{-}corr} = 0.007$) in the NAcc in the right hemisphere. (Left NAcc activation was not revealed as significant in this analysis for either δ_c or δ_e). This suggests that the phasic dopaminergic response encoded the egocentric prediction error.

Activity significantly correlated with δ_e (see Fig. 4 and Supplementary Table 6) was found in regions involved in response inhibition (hypothesised on the basis of their previous activation in tasks involving response inhibition — see Discussion): IFGr, MFGr, cingulate gyrus, globus pallidus pars externa, bilateral premotor cortex, left insula, left inferior parietal lobule, right and left precuneus, and left lateral orbitofrontal cortex (OFC). Unhypothesised activity (which survived a more conservative threshold of p<0.0001 uncorrected) was noted in right frontopolar cortex (FPC), left cuneus and right parahippocampal gyrus.

Activity related to player's exploration

We sought activity in frontal regions and identified two clusters of activity with peaks in IFGr and dorso-lateral prefrontal cortex (DLPFC) (Supplementary Table 7). Unhypothesised activity (surviving a more conservative threshold of p<0.0001 uncorrected) was also observed in three other regions: right precuneus, the substantia nigra and subthalamic nucleus.

Discussion

Representation of the competitor's actions

When players observed their competitor's bandit selection, they activated premotor and motor cortex in regions activated when making the same selection themselves. Apart from supporting our proposal for the role of action representation in competitive learning, it is also notable that activation of these regions of the MNS occurred without seeing any biological movement, and in response to decisions that players knew were computer generated. This was predicted on the basis of studies showing motor activation when actions are suggested in ways other than through direct visual observation of a movement. For example, it has been shown that primates activate MNS when hearing the noise associated with an action but without seeing it (Kohler et al., 2002). In our study, movement was suggested only by the change in color of a bandit. However, it has been demonstrated that static pictures suggestive of actions are sufficient to generate motor activity in the observer when the goals of the represented actions are understood, which would have been the case here (Johnson-Frey et al., 2003). The fact that the stimulus exciting motor activity was a signal from a computer, representing an action which the player knows is not being biologically executed is, perhaps, more surprising. Research has shown that an action such as biting can produce MNS activity whether performed by human or dog, but not barking — demonstrating that the motor system can be excited by an observed action provided by another species, if it is in the repertoire of the observer (Buccino et al., 2004). Our results demonstrate that outcomes merely suggesting the virtual actions of an artificial agent can also activate the mirror neuron system. In our task, the player representing their own and the competitor's actions in a similar way may support learning, by helping the player associate actions with outcomes, irrespective of who performed them.

The MNS also includes the inferior parietal lobule and the IFG (Rizzolatti and Craighero, 2004). Activity in both these regions, however, when individuals were observing grasping actions, has been revealed not to be hand-specific (Shmuelof and Zohary, 2006), prompting suggestions of their possible sensitivity to aspects of the task such as its meaning, complexity or relation to the observer's task.

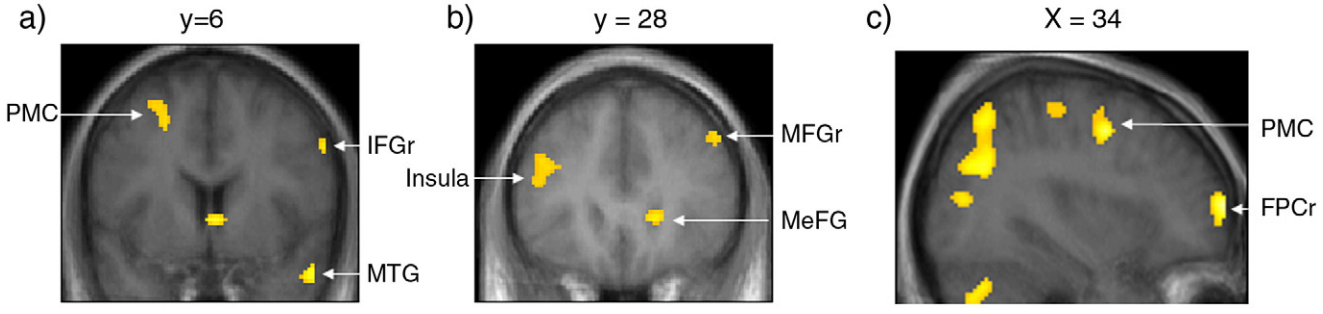

Fig. 4. Activity correlated with the learning signal (δ_e — egocentric prediction error) when the player is observing the competitor's feedback (p<0.001 uncorrected): a) Coronal section (y = 6), showing activation in right inferior frontal gyrus (IFGr, centre at x = 60, y = 7, z = 9), as well as premotor cortex (PMC) and middle temporal gyrus (MTG) b) Coronal section (y = 28) showing activation in right middle frontal gyrus (MFGr, centre at x = 48, y = ,29 z = 32) as well as medial frontal gyrus (MeFG) and left insula c) Sagittal section (x = 34) showing the activation in right frontopolar cortex (FPC, centre at x = 34, y = 62, z = −1) and other regions including right premotor cortex (PMC).

Hence, we were not expecting these regions to be activated in an analysis based on handedness.

Activity related with magnitude of competitor's reward

Positive correlation of activity with the points won by the competitor did not reach threshold in any region of the brain, while negatively correlated activity was revealed in the posterior medial frontal cortex. The results of de Bruijn et al. (2009), who studied activations in response to the monitoring of errors by participants and by their competitors/collaborators, showed this region to be active regardless of who was making the errors. In our study, activity in this region was not significantly negatively correlated with the players' own magnitude of reward suggesting that, compared with their own actions, players were monitoring the failure rather than the success of their competitor.

The medial frontal cortex has been previously implicated in reinforcement learning in three different ways (Rushworth, 2008; Rushworth and Behrens, 2008). Firstly, it is thought to be involved with the updating of action values and in mediating the impact of past reinforcement history on the next choice made, so influencing learning rate. Secondly, this region is thought important when an exploratory action is generated and, thirdly, it is considered critical when conflicting information in the immediate environment instructs more than one possible response. In our study, despite the outcomes of the competitor's choices always being positive in absolute value, the player appears to be learning from the extent to which these actions may be judged as errors (i.e. the extent to which the competitor's outcomes are less than expected). Some "errors" by the competitor in our study might be expected to carry increased amounts of information that would justify a greater rate of information update. For example, when the competitor makes a random and low-scoring exploratory selection of a rarely sampled bandit, this rare glimpse of the bandit's content might justify an increase in learning rate. However, the pMFC activation observed in this analysis is in response to poor outcomes of the competitor's actions in absolute terms, irrespective of whether these provide new information or, instead, are quite predictable. This suggests it may also be attributable to increased conflict and the need to consider multiple and exploratory responses, as when a previously rich bandit begins to provide pay outs that become predictably low.

Inhibition and learning from competitors

Many regions where activity was correlated with δ_e were those typically activated by tasks involving response inhibition: IFGr (Aron et al., 2007; Aron and Poldrack, 2005, 2006; Cai and Leung, 2009; Coxon et al., 2009; Garavan et al., 2006; Garavan et al., 1999; Konishi et al., 1998; Leung and Cai, 2007; Pliszka et al., 2006; Rubia et al., 2001; Rubia et al., 2003), cingulate gyrus (Garavan et al., 2006; Garavan et al., 1999; Luna, 2004; Rubia et al., 2001; Rubia et al., 2003), MFGr (Aron and Poldrack, 2005, 2006; Cai and Leung, 2009; Coxon et al., 2009; Garavan et al., 2006; Garavan et al., 1999; Konishi et al., 1998; Pliszka et al., 2006; Rubia et al., 2001), left insula (Garavan et al., 2006; Leung and Cai, 2007), left inferior parietal lobule (Aron et al., 2007; Garavan et al., 2006, 1999; Luna, 2004; Rubia et al., 2001; Rubia et al., 2003), premotor cortex (Cai and Leung, 2009; Leung and Cai, 2007), right (Garavan et al., 2006) and left precuneus(Cai and Leung, 2009; Rubia et al., 2001) and globus pallidus pars externa (Frank et al., 2004).

We propose two possible interpretations of this apparent correlation of egocentric prediction error with activity in regions associated with response inhibition (that are not mutually exclusive). First, we propose that our results may support a model in which the brain's response inhibition system is critically involved in learning from the competitor's feedback. More specifically, if it becomes apparent that the competitor's action has led to the competitor's loss, the inhibition of this action is strengthened. In this model, the component of value learnt from competitor's feedback is encoded separately from the component of value learnt from the player's own feedback, hence we refer to this model as the *dual-value* model. In the dual-value model, after competitor's feedback, m_i are *not* modified according to Eq. (2), but instead the strengths of synaptic connections to regions involved in action inhibition are modified proportionally to the level of dopamine. Let us denote the inhibition of action i by n_i, and assume that the probability of selecting action i depends on $m_i - n_i$ (i.e. the larger n_i, the lower the probability of selecting action i). In the dual-value model, n_i are modified proportionally to the egocentric prediction error, so that if the competitor's action produces a loss, it will be inhibited more in the future.

$$n_i \leftarrow n_i + \eta\delta_e \tag{8}$$

Since in humans the dopaminergic neurons project not only to the striatum but to almost entire cerebral cortex (Camps et al., 1989; Cortes et al., 1989), in the dual-value model, n_i are encoded in the strengths of synaptic connections to regions involved in action inhibition. The advantage of this model is that the synaptic weights encoding n_i are modified depending on the level of dopamine in a similar manner as in the striatum (compare Eqs. (1) and (8)) and the dual-value model relies on a known mechanism of synaptic plasticity. Thus, the model predicts neural activity correlated with δ_e in regions related to response inhibition, as we observed in the experiment.

In the Supplementary content, we show that the dual-value model predicts the same behavior as the model described in the Introduction, and hence these two models are indistinguishable on the basis of our behavioral data. Furthermore, the dual-value model generates exactly the same values of prediction errors as the model described in the Introduction, hence if the dual-value model were used to generate regressors for imaging analysis (listed in Table 2), the results of the analysis would not change.

In the dual-value model, the components of value learnt from own and competitor's feedback are encoded by two separate sets of synapses. Analogous separate encoding of components of value was demonstrated by Frank et al. (2004) who provided evidence that the components of value learnt from positive and negative feedback are stored in synapses of separate striatal neurons that then jointly influence the choice. In the present context, the dual-value model would predict that, at the time of player's choice, separate brain regions should have activity correlated with m_i and n_i (Peter Dayan, personal communication). Future experimental paradigms may succeed in disentangling players' responses to the competitor's feedback and their subsequent decision, allowing such a prediction to be tested.

The second possible interpretation of the correlation between δ_e and the activity of the response inhibition system is that this system inhibits an automatic tendency for imitation of the competitor's action. It would appear that processes linked to representing the competitor's actions begin prior to the outcomes of a competitor's decision being revealed, and that these processes are subsequently inhibited when the competitor's actions have unfavourable results. There is a widespread view that the MNS is crucially involved in imitation (Brass and Heyes, 2005; Buccino et al., 2004; Heyes, 2001; Iacoboni, 2005; Iacoboni and Dapretto, 2006; Rizzolatti, 2005; Rizzolatti et al., 2001) and this neural representation of the competitor's actions may form a critical part the player's preparation to imitate. However, an alternative conceptualisation of the MNS might restrict the role of such representation to processes merely attributing unexpected loss to a particular action. The exact role of the neural representation of the competitor's actions in the player's brain cannot be determined from the present study, but would be an interesting area for future research.

Previous reports of a fronto-parietal network being involved with movement inhibition have involved tasks with an explicit cue triggering inhibitory control (e.g. the Go–No Go task). However, both of the above interpretations of our data seem to suggest that a similar network may be involved in inhibitory control in tasks where only implicit information is provided (e.g. in our task, a poor outcome of the competitor's action).

Other activity related with competitor's feedback

We have reported that the regions correlated with egocentric prediction error included lateral OFC and FPC. Below we discuss relationships between our results and results of other studies that reported activity in these two regions.

The lateral OFC is thought to subserve the suppression of previously rewarded responses in uncertain situations (Elliott and Dolan, 1999; Iversen and Mishkin, 1970). In our task, we recorded left lateral OFC activation correlated with δ_e. That is, OFC activation increased with the degree of unexpectedly poor outcomes for the competitor, as the desirability of suppressing previously rewarded responses from the selected bandit might be expected to increase. At the same time, however, unexpectedly low outcomes for the competitor increase the likelihood that the player must plan an exploratory choice, and the posterior region of the OFC activated in our analysis has also been implicated in the excitement generated by risky choices(Elliott et al., 2000).

We also observed a cluster with activity correlated with δ_e in right FPC and numerous studies that suggest FPC has a critical role when human players switch between tasks, particularly when it is necessary to hold information about one task in working memory (Braver et al., 2003; Koechlin et al., 1999; Koechlin and Hyafil, 2007; Koechlin and Summerfield, 2007; Ramnani and Owen, 2004). The unexpected failure of a competitor's action would likely increase the player's consideration of alternatives and could, therefore, be expected to increase activation of this region. More specifically, a recent fMRI study involving a 2-armed bandit task supports the role of FPC in coding the value of such alternatives, in terms of their relative advantage (Boorman et al., 2009). Our findings are also aligned with this proposed role for FPC, since the relative value of all bandits alternative to the one selected increases with positive δ_e. On this basis, the neural correlates of egocentric prediction error in this region may relate to individual differences in players' tendency to seek alternative selections to those made by the competitor. To explore such a relation (see Supplementary Fig. 2), we carried out a post-hoc analysis and found δ_e related activity in FPC was positively related to alternative behavior ($r = 0.53$, $p = 0.034$).

We also found activity correlated with δ_e in regions likely to be involved with visual processing of stimuli (BA 17,18) and encoding of information relating to outcomes (BA 35, parahippocampal gyrus (Liu and Richmond, 2000)). We also observed activation of the cerebellum which is not considered directly involved with response inhibition (Thoma et al., 2008) but has been linked to more global preparation associated with smooth and optimal performance of subsequent motor and cognitive functions (Courchesne and Allen, 1997).

Activity related to player's exploration

A previous study (Daw et al., 2006) had employed a 4-armed bandit task without a competitor. In that study, FPC and intraparietal sulcus were preferentially active during the player's exploratory decisions. In our study, as expected, activity associated with exploration was observed in frontal regions although, rather than in FPC as observed previously, this activity was more caudally situated with peaks in IFGr and dorso-lateral prefrontal cortex (DLPFC). The difference in the pattern of exploration-related activity between this previous study and our own may arise from the competitor influencing the player's selection strategy. In our experiment only 17% of players' selections were exploratory whereas, when a similar task was played without a competitor in the study of Daw et al., 36% of choices were exploratory. In our task, in addition to their own exploratory selections, the player could gain information about bandit values from the competitors' exploratory selections. When available, this was a safer source of information for players than their own exploratory decisions, which might result in sampling a low-yielding bandit. (Indeed, there was a negative correlation across participants between the proportion of their selections that were exploratory and the total reward they earned during the game, $r = -0.74$, $p < 0.001$). In post-task interviews, players spoke of the helpfulness of observing the competitor's exploratory behavior, and commented that they sometimes allowed their competitor to "take the risk" of additional information gathering.

Activity was also observed in the substantia nigra and subthalamic nucleus, suggesting activation of an inhibitory network of IFGr, the substantia nigra and subthalamic nucleus (Aron et al., 2007; Aron and Poldrack, 2006). The subthalamic nucleus provides non-specific inhibition of all motor programs (Mink, 1996) and there are two possible explanations of its activity on exploratory trials. Firstly, the subthalamic nucleus has been proposed to provide inhibition proportional to the level of conflict in evidence supporting different alternatives (Bogacz and Gurney, 2007; Frank, 2006), and one could argue that players experienced high conflict on exploratory trials. However, it is unlikely that this function of the subthalamic nucleus was manifested in our experiment, because: (i) Other regions associated with conflict (e.g. the anterior cingulate cortex (MacDonald et al., 2000)) were not significantly more active on exploratory trials. (ii) The subthalamic nucleus has been shown to be critical in decisions between highly rewarding options rather than poor options (Frank et al., 2007) but, in our study, players' exploratory choices tended to occur after trials in which they had received lower rewards: The average player's reward on their trials preceding their exploratory selections ($r_p = 59.8$) was lower than those preceding their exploitative selections ($r_p = 71.5$). Secondly, it has been shown that when the current motor program is stopped, the subthalamic nucleus becomes active due to the input it receives from cortical regions involved in action inhibition including IFGr (Aron and Poldrack, 2006). It is likely that the subthalamic nucleus provided such inhibition on exploratory trials because: (i) IFGr had increased activity on exploratory trials. (ii) In order to make exploratory selection it might be necessary to first block the current motor program of selecting action with highest m_i.

If the response inhibition system is involved in both learning from competitors and exploration, then one could expect that the activation in the response inhibition network in the competitor's turn should predict the exploratory behavior in the player's subsequent turn. Indeed, we observed that egocentric prediction error was significantly higher before players' exploratory selections than exploitative selections (unpaired t-test $p < 10^{-6}$).

We also observed increased DLPFC activity for players' exploratory trials. This might reflect increased working memory demands associated with making an exploratory decision in a competitive environment, with neurons in this region encoding past decisions and payoffs for primates in a competitive game (Barraclough et al., 2004). The e-greedy competitor would frequently have sampled bandits that the participant might not have explored themselves. These types of exploratory decision by the competitor might not have obviated the need for the player to make their own exploratory decisions, but would have increased the amount of information available when doing so. The sharing of risk-taking, and the associated tendency of the player to make exploratory decisions after low rewards, may also help explain why we did not observe activity in FPC on exploratory trials, as in the study by Daw et al. In their study, FPC activation on exploratory trials was explained in terms of the additional cognitive control required to override default exploitative tendencies. However, in our task, the exploitative tendency on exploratory trials would be less if preceding rewards had been lower, and so require less cognitive control to override them (Peter Dayan, personal communication).

In conclusion, our results support a model of competitive reward learning in which the player generates neural representations of their competitor's actions prior to outcomes becoming known, possibly in readiness to initiate these actions. At the outcome of the competitor's selection, we observed activities suggesting reward-based response inhibition and the appraisal of alternatives, in relation to a learning signal provided by the competitor's unexpected losses. In competitive foraging, processes involving the mirror neuron, response inhibition and reward systems may cooperate in supporting efficient reward exploitation and loss avoidance.

Acknowledgments

We thank Matthew Rushworth and Sander Nieuwenhuis for reading an earlier version of the manuscript and for their very useful comments, and Peter Dayan and Cecilia Heyes for discussion and very useful advice regarding data analysis. We are also grateful for the technical support provided by Ian Summers and Abdelmalek Benattayallah at the Peninsula MR Research Centre (Exeter), and staff at the computational facilities of the Advanced Computing Research Centre (University of Bristol).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2010.06.027.

References

Akaike, H., 1974. New look at statistical-model identification. Ieee Trans. Autom. Control 19, 716–723.
Andersson, J.L.R., Hutton, C., Ashburner, J., Turner, R., Friston, K., 2001. Modelling geometric deformations in EPI time series. Neuroimage 13, 903–919.
Aron, A.R., Poldrack, R.A., 2005. The cognitive neuroscience of response inhibition: Relevance for genetic research in attention-deficit/hyperactivity disorder. Biol. Psychiatry 57, 1285–1292.
Aron, A.R., Poldrack, R.A., 2006. Cortical and subcortical contributions to stop signal response inhibition: role of the subthalamic nucleus. J. Neurosci. 26, 2424–2433.
Aron, A.R., Behrens, T.E., Smith, S., Frank, M.J., Poldrack, R.A., 2007. Triangulating a cognitive control network using diffusion-weighted magnetic resonance imaging (MRI) and functional MRI. J. Neurosci. 27, 3743–3752.
Barraclough, D.J., Conroy, M.L., Lee, D., 2004. Prefrontal cortex and decision making in a mixed-strategy game. Nat. Neurosci. 7, 404–410.
Bogacz, R., Gurney, K., 2007. The basal ganglia and cortex implement optimal decision making between alternative actions. Neural Comput. 19, 442–477.
Boorman, E.D., Behrens, T.E.J., Woolrich, M.W., Rushworth, M.F.S., 2009. How green is the grass on the other side? Frontopolar cortex and the evidence in favor of alternative courses of action. Neuron 62, 733–743.
Brass, M., Heyes, C., 2005. Imitation: is cognitive neuroscience solving the correspondence problem? [Review]. Trends Cogn. Sci. 9 (10), 489–495.
Braver, T.S., Reynolds, J.R., Donaldson, D.I., 2003. Neural mechanisms of transient and sustained cognitive control during task switching. Neuron 39, 713–726.
Buccino, G., Lui, F., Canessa, N., Patteri, I., Lagravinese, G., Benuzzi, F., Porro, C.A., Rizzolatti, G., 2004. Neural circuits involved in the recognition of actions performed by nonconspecifics: an fMRI study. J. Cogn. Neurosci. 16, 114–126.
Caetano, G., Jousmaki, V., Hari, R., 2007. Actor's and observer's primary motor cortices stabilize similarly after seen or heard motor actions. Proc. Natl. Acad. Sci. U. S. A. 104 (21), 9058–9062.
Cai, W.D., Leung, H.C., 2009. Cortical activity during manual response inhibition guided by color and orientation cues. Brain Res. 1261, 20–28.
Camps, M., Cortes, R., Gueye, B., Probst, A., Palacios, J.M., 1989. Dopamine-receptors in human-brain — autoradiographic distribution of D2 sites. Neuroscience 28, 275–290.
Cochin, S., Barthelemy, C., Roux, S., Martineau, J., 1999. Observation and execution of movement: similarities demonstrated by quantified electroencephalography. [Article]. Eur. J. Neurosci. 11 (5), 1839–1842.
Coxon, J.P., Stinear, C.M., Byblow, W.D., 2009. Stop and go: the neural basis of selective movement prevention. [Article]. J. Cogn. Neurosci. 21 (6), 1193–1203.
Cortes, R., Gueye, B., Pazos, A., Probst, A., Palacios, J.M., 1989. Dopamine-receptors in human-brain — autoradiographic distribiution of D1 sites. Neuroscience 28, 263–273.
Courchesne, E., Allen, G., 1997. Prediction and preparation, fundamental functions of the cerebellum. Learn. Mem. 4, 1–35.
D'Ardenne, K., McClure, S.M., Nystrom, L.E., Cohen, J.D., 2008. BOLD responses reflecting dopaminergic signals in the human ventral tegmental area. [Article]. Science 319 (5867), 1264–1267.
Daw, N., O'Doherty, J.P., Dayan, P., Seymour, B., Dolan, R.J., 2006. Cortical substrates for exploratory decisions in humans. Nature 441, 876–879.
Dayan, P., Abbott, L.F., 2001. Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems. MIT Press, Cambridge, MA.
de Bruijn, E.R.A., de Lange, F.P., von Cramon, D.Y., Ullsperger, M., 2009. When errors are rewarding. J. Neurosci. 29, 12183–12186.
Duvernoy, H.M., 1999. The Human Brain. Springer-Verlag, Vienna.
Elliott, R., Dolan, R.J., 1999. Differential neural responses during performance of matching and nonmatching to sample tasks at two delay intervals. J. Neurosci. 19, 5066–5073.
Elliott, R., Dolan, R.J., Frith, C.D., 2000. Dissociable functions in the medial and lateral orbitofrontal cortex: evidence from human neuroimaging studies. Cereb. Cortex 10, 308–317.
Forman, S.D., Cohen, J.D., Fitzgerald, M., Eddy, W.F., Mintun, M.A., Noll, D.C., 1995. Improved assessment of significant activation in functional magnetic-resonance-imaging (fMRI) — use of a cluster-size threshold. Magn. Reson. Med. 33, 636–647.
Frank, M.J., 2006. Hold your horses: a dynamic computational role for the subthalamic nucleus in decision making. Neural Netw. 19, 1120–1136.
Frank, M.J., Seeberger, L.C., O'Reilly, R.C., 2004. By carrot or by stick: cognitive reinforcement learning in Parkinsonism. Science 306, 1940–1943.
Frank, M.J., Samanta, J., Moustafa, A.A., Sherman, S.J., 2007. Hold your horses: impulsivity, deep brain stimulation, and medication in parkinsonism. Science 318, 1309–1312.
Friston, K.J., Penny, W.D., Glaser, D.E., 2005. Conjunction revisited. Neuroimage 25, 661–667.
Garavan, H., Ross, T.J., Stein, E.A., 1999. Right hemispheric dominance of inhibitory control: an event-related functional MRI study. Proc. Natl Acad. Sci. USA 96, 8301–8306.
Garavan, H., Hester, R., Murphy, K., Fassbender, C., Kelly, C., 2006. Individual differences in the functional neuroanatomy of inhibitory control. Brain Res. 1105, 130–142.
Groos, K., 1898. The Play of Animals. Appleton, New York.
Hari, R., Forss, N., Avikainen, S., Kirveskari, E., Salenius, S., Rizzolatti, G., 1998. Activation of human primary motor cortex during action observation: a neuromagnetic study [Article]. Proc. Natl. Acad. Sci. U. S. A. 95 (25), 15061–15065.
Heyes, C., 2001. Causes and consequences of imitation. [Review]. Trends Cogn. Sci. 5 (6), 253–261.
Iacoboni, M., 2005. Neural mechanisms of imitation. [Article]. Curr. Opin. Neurobiol. 15 (6), 632–637.
Iacoboni, M., Dapretto, M., 2006. The mirror neuron system and the consequences of its dysfunction [Review]. Nat. Rev. Neurosci. 7 (12), 942–951.
Iversen, S.D., Mishkin, M., 1970. Perseverative interference in monkeys following selective lesions of inferior prefrontal convexity. Exp. Brain Res. 11, 376.
Johnson-Frey, S.H., Maloof, F.R., Newman-Norlund, R., Farrer, C., Inati, S., Grafton, S.T., 2003. Actions or hand–object interactions? Human inferior frontal cortex and action observation. Neuron 39, 1053–1058.
Koechlin, E., Hyafil, A., 2007. Anterior prefrontal function and the limits of human decision-making. Science 318, 594–598.
Koechlin, E., Summerfield, C., 2007. An information theoretical approach to prefrontal executive function. Trends Cogn. Sci. 11, 229–235.
Koechlin, E., Basso, G., Pietrini, P., Panzer, S., Grafman, J., 1999. The role of the anterior prefrontal cortex in human cognition. Nature 399, 148–151.
Kohler, E., Keysers, C., Umilta, M.A., Fogassi, L., Gallese, V., Rizzolatti, G., 2002. Hearing sounds, understanding actions: action representation in mirror neurons. Science 297, 846–848.
Konishi, S., Nakajima, K., Uchida, I., Sekihara, K., Miyashita, Y., 1998. No-go dominant brain activity in human inferior prefrontal cortex revealed by functional magnetic resonance imaging. [Article]. Eur. J. Neurosci. 10 (3), 1209–1213.
Leung, H.C., Cai, W.D., 2007. Common and differential ventrolateral prefrontal activity during inhibition of hand and eye movements. J. Neurosci. 27, 9893–9900.
Liu, Z., Richmond, B.J., 2000. Response differences in monkey TE and perirhinal cortex: stimulus association related to reward schedules. J. Neurophysiol. 83, 1677–1692.
Luna, B., 2004. Algebra and the adolescent brain. Trends Cogn. Sci. 8, 437–439.
MacDonald, A.W., Cohen, J.D., Stenger, V.A., Carter, C.S., 2000. Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. Science 288, 1835–1838.
McClure, S.M., Berns, G.S., Montague, P.R., 2003. Temporal prediction errors in a passive learning task activate human striatum. [Article]. Neuron 38 (2), 339–346.
Mikl, M., Marecek, R., Hlustik, P., Pavlicova, M., Drastich, A., Chlebus, P., et al., 2008. Effects of spatial smoothing on fMRI group inferences. Magn. Reson. Imaging 26 (4), 490–503.
Miller, E.K., Cohen, J.D., 2001. An integrative theory of prefrontal cortex function. Annu. Rev. Neurosci. 24, 167–202.
Mink, J.W., 1996. The basal ganglia: focused selection and inhibition of competing motor programs. Prog. Neurobiol. 50, 381–425.
Montague, P.R., Dayan, P., Sejnowski, T.J., 1996. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J. Neurosci. 16, 1936–1947.
Nelder, J.A., Mead, R., 1965. A simplex method for function minimization. Comput. J. 7, 308–313.
Neto, L.L., Oliveira, E., Correia, F., Ferreira, A.G., 2008. The human nucleus accumbens: where is it? A stereotactic, anatomical and magnetic resonance imaging study. Neuromodulation 11, 13–22.
O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., Dolan, R.J., 2004. Dissociable roles of ventral and dorsal striatum in instrumental conditioning. [Article]. Science 304 (5669), 452–454.
Pliszka, S.R., Glahn, D.C., Semrud-Clikeman, M., Franklin, C., Perez, R., Xiong, J.J., 2006. Neuroimaging of inhibitory control areas in children with attention deficit hyperactivity disorder who were treatment naive or in long-term treatment. Am. J. Psychiatry 163, 1052–1060.
Ramnani, N., Owen, A.M., 2004. Anterior prefrontal cortex: insights into function from anatomy and neuroimaging. Nat. Rev. Neurosci. 5, 184–194.
Reynolds, J.N.J., Wickens, J.R., 2002. Dopamine-dependent plasticity of corticostriatal synapses. Neural Netw. 15, 507–521.

Reynolds, J.N.J., Hyland, B.I., Wickens, J.R., 2001. A cellular mechanism of reward-related learning. Nature 413, 67–70.
Rizzolatti, G., 2005. The mirror neuron system and its function in humans. Anat. Embryol. 210 (5–6), 419–421.
Rizzolatti, G., Craighero, L., 2004. The mirror neuron system. Annu. Rev. Neurosci. 27, 169–192.
Rizzolatti, G., Fogassi, L., Gallese, V., 2001. Neurophysiological mechanisms underlying the understanding and imitation of action. [Editorial Material]. Nat. Rev. Neurosci. 2 (9), 661–670.
Rubia, K., Russell, T., Overmeyer, S., Brammer, M.J., Bullmore, E.T., Sharma, T., Simmons, A., Williams, S.C.R., Giampietro, V., Andrew, C.M., Taylor, E., 2001. Mapping motor inhibition: conjunctive brain activations across different versions of go/no-go and stop tasks. Neuroimage 13, 250–261.
Rubia, K., Smith, A.B., Brammer, M.J., Taylor, E., 2003. Right inferior prefrontal cortex mediates response inhibition while mesial prefrontal cortex is responsible for error detection. [Article]. Neuroimage 20 (1), 351–358.
Rushworth, M.F.S., 2008. Intention, choice, and the medial frontal cortex. In Year in Cognitive Neuroscience 2008, vol. 1124. Blackwell Publishing, Oxford, pp. 181–207.
Rushworth, M.F.S., Behrens, T.E.J., 2008. Choice, uncertainty and value in prefrontal and cingulate cortex. [Review]. Nat. Neurosci. 11 (4), 389–397.
Schultz, W., Dayan, P., Montague, P.R., 1997. A neural substrate of prediction and reward. [Article]. Science 275 (5306), 1593–1599.
Shmuelof, L., Zohary, E., 2006. A mirror representation of others' actions in the human anterior parietal cortex. J. Neurosci. 26, 9736–9742.
Sutton, R.S., Barto, A.G., 1998. Reinforcement Learning: An Introduction. MIT Press, Cambridge, MA.
Thoma, P., Koch, B., Heyder, K., Schwarz, M., Daum, I., 2008. Subcortical contributions to multitasking and response inhibition. [Article]. Behav. Brain Res. 194 (2), 214–222.
Tobler, P.N., Fiorillo, C.D., Schultz, W., 2005. Adaptive coding of reward value by dopamine neurons. Science 307 (5715), 1642–1645.
Touzalin-Chretien, P., Dufour, A., 2008. Motor cortex activation induced by a mirror: evidence from lateralized readiness potentials. J. Neurophysiol. 100, 19–23.
Zaghloul, K.A., Blanco, J.A., Weidemann, C.T., McGill, K., Jaggi, J.L., Baltuch, G.H., et al., 2009. Human Substantia Nigra Neurons Encode Unexpected Financial Rewards. [Article]. Science 323 (5920), 1496–1499.