EJN EUROPEAN JOURNAL OF NEUROSCIENCE FENS Federation of European Neuroscience Societies

European Journal of Neuroscience, Vol. 41, pp. 1358–1371, 2015 doi:10.1111/ejn.12897

COGNITIVE NEUROSCIENCE

Differentiating neural systems mediating the acquisition vs. expression of goal-directed and habitual behavioral control

EJN FEATURED ARTICLE

Mimi Liljeholm,[1,2] Simon Dunne[1,3] and John P. O'Doherty[1,3]

[1]Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland

[2]Department of Cognitive Sciences, University of California, 2312 Social & Behavioral Science Gateway Building, Irvine, CA 92697, USA

[3]Division of the Humanities and Social Sciences and Computation and Neural Systems Program, California Institute of Technology, Pasadena, CA, USA

Keywords: action–outcome, functional magnetic resonance imaging, habit formation, human, stimulus–response

Abstract

Considerable behavioral data indicate that operant actions can become habitual, as demonstrated by insensitivity to changes in the action–outcome contingency and in subjective outcome values. Notably, although several studies have investigated the neural substrates of habits, none has clearly differentiated the areas of the human brain that support habit formation from those that implement habitual control. We scanned participants with functional magnetic resonance imaging as they learned and performed an operant task in which the conditional structure of the environment encouraged either goal-directed encoding of the consequences of actions, or a habit-like mapping of actions to antecedent cues. Participants were also scanned during a subsequent assessment of insensitivity to outcome devaluation. We identified dissociable roles of the cerebellum and ventral striatum, across learning and test performance, in behavioral insensitivity to outcome devaluation. We also showed that the inferior parietal lobule (an area previously implicated in several aspects of goal-directed action selection, including the attribution of intent and awareness of agency) predicted sensitivity to outcome devaluation. Finally, we revealed a potential functional homology between the human subgenual cortex and rodent infralimbic cortex in the implementation of habitual control. In summary, our findings suggested a broad systems division, at the cortical and subcortical levels, between brain areas mediating the encoding and expression of action–outcome and stimulus–response associations.

Introduction

Habitual action selection is defined by insensitivity to changes in the causal efficacy with which actions produce rewards and to the current subjective value of those rewards (Balleine & Dickinson, 1998). Neuroscientific research on humans and rodents has demonstrated that the brain areas mediating habitual performance are dissociable from those supporting more deliberate, goal-directed, action selection (Balleine & Dickinson, 1998; Yin *et al.*, 2004, 2005; Valentin *et al.*, 2007; Tricomi *et al.*, 2009). Intriguingly, work in rodents also suggests that distinct neural substrates make specialized contributions to the development vs. deployment of habits (Killcross & Coutureau, 2003). In contrast, in humans there has been no clear differentiation between brain areas that support habit formation and those that implement habitual control. Although a couple of neuroimaging studies have demonstrated increases in neural activity concomitant with the development of habits in a posterior area of the lateral striatum (Tricomi *et al.*, 2009; Wunderlich *et al.*, 2012), the use of overtraining to induce habitual responding in these studies confounds well-trained performance with habitual control. In the current study, in order to discriminate between neural substrates supporting the acquisition vs. expression of habits, we scanned human participants with functional magnetic resonance imaging as they performed a novel instrumental task (see Task section in Materials and methods), designed to rapidly induce habitual responding without the potentially confounding process of overtraining.

Pharmacological disruptions and electrophysiological recordings of the rodent brain have strongly implicated the dorsolateral striatum in the acquisition of habits. Pre-training lesions of the dorsolateral striatum abolish habit formation (Yin *et al.*, 2004), and distinct changes in neuronal activity patterns (Jog *et al.*, 1999), including substantial decreases in firing rates (Carelli *et al.*, 1997; Tang *et al.*, 2007), have been demonstrated in this area across the development of habitual responding. In contrast, the infralimbic region of the prefrontal cortex has been suggested to support an executive control system that facilitates the expression of habits. Post-training muscimol-induced inactivation of the infralimbic cortex disrupts habitual performance (Coutureau & Killcross, 2003; Haddon & Killcross, 2011), and changes in neuronal ensemble activity patterns in this

Correspondence: Mimi Liljeholm, [2]Department of Cognitive Sciences, as above.
E-mail: m.liljeholm@uci.edu

Received 3 October 2014, revised 6 March 2015, accepted 12 March 2015

© 2015 Federation of European Neuroscience Societies and John Wiley & Sons Ltd

area occur very late in training and closely track the behavioral manifestation of habitual control (Smith & Graybiel, 2013). Further evidence for the involvement of the infralimbic cortex in the expression of habits comes from studies in which optogenetic perturbation of this area disrupts well-ingrained habitual behavior (Smith *et al.*, 2012). Based on these findings in rodents, we hypothesized that human homologs of the dorsolateral striatum and infralimbic cortex would be involved in the formation and expression of habits respectively, such that behavioral insensitivity to outcome devaluation would correlate with neural activity during learning in the dorsal putamen (Carelli & West, 1991; Draganski *et al.*, 2008), but with activity during actual test performance in the subgenual cortex (Ongur & Price, 2000; Ongur *et al.*, 2003).

Materials and methods

Participants

Twenty volunteers (mean age 21.4 ± 2.63 years, range 19–28 years; 11 males) participated in the experiment. Due to technical problems (loss of power to the stimulus computer), one of the subjects was excluded, yielding a total of 19 participants. All participants were healthy and were recruited locally from the city of Dublin, Ireland. The study was approved by the Trinity College Dublin School of Psychology research ethics committee, and all participants gave informed consent. The study conformed to the guidelines set out in the 2013 World Medical Association Declaration of Helsinki.

Task

Goal-directed actions, defined by their sensitivity to changes in both action–outcome contingency and outcome value, have been proposed to depend on an internal model of the world that explicitly relates alternative actions to future environmental states (Doya *et al.*, 2002; Daw *et al.*, 2005). Consistent with this theoretical framework, data from rodent studies suggest that reliance on a goal-directed vs. habitual strategy might depend on the ease with which alternative actions can be associated with distinct outcomes; goal-directed performance appears to dominate, in spite of overtraining, when alternative actions yield distinct sensory-specific outcomes (e.g. grain vs. sucrose pellets) (Colwill & Rescorla, 1985; Holland, 2004), as well as when the rate of outcome delivery depends on the rate of responding, rather than on a particular time interval passing between successive reinforced responses (i.e. ratio vs. interval schedules of reinforcement) (Dickinson *et al.*, 1983). The current task was structured on these potential bases of behavioral control, in that external contingencies either facilitated or impeded a reliable mapping of alternative actions to distinct sensory-specific outcome states.

Participants were required to maintain the balance of a system of fluid-filled beakers (see Fig. 1A for details). As long as all beakers had sufficient fluid, system balance was maintained and randomly occurring balance checks yielded monetary reward. However, on each trial, one of the beakers would be emptied causing ‘system imbalance’, with balance checks resulting in monetary loss until the participant refilled the beaker by performing a particular instrumental action. The emptying of a beaker was always accompanied by the onset of one of four abstract cues. In the Stimulus–Response (S-R) condition, the identity of the presented cue determined which instrumental action would refill the emptied beaker, regardless of which of the beakers had lost its fluid. Consequently, across trials, each rewarded action was paired with a specific antecedent cue, but was decorrelated from the refilling of any particular beaker. Conversely, in a Response–Outcome (R-O) condition, each instrumental action refilled a particular beaker, regardless of which abstract cue was presented, such that identification of the relevant subgoal (e.g. refilling beaker 1), combined with knowledge about specific action–outcome contingencies (i.e. action 1 refills beaker 1), indicated which action would restore system balance. To ensure that discriminatory neural activity was not due to differences in the visual processing of abstract cues vs. beakers, a matching task was interleaved with the instrumental task (see Fig. 1B).

The persistent execution of an action after its outcome has been devalued is a defining feature of habitual performance (Adams & Dickinson, 1981; Adams, 1982). In the current study, following acquisition of the instrumental task, we devalued one of the four beakers by degrading its relationship to monetary gain, such that system balance was maintained, and continued to yield points even when the liquid in this beaker dropped below threshold. Because there was a small cost for regulating the system, attempts to refill this beaker now resulted in a net loss (see Fig. 1A and Experimental Procedure section for details). Based on the notion that failure to associate alternative actions with distinct outcome states obstructs goal-directed encoding, we predicted that the S-R condition would bolster habit formation during acquisition and bias participants towards habitual action selection, defined as responding to refill the devalued beaker, in subsequent test performance. Note that, in both conditions, distinct features of the stimulus environment can enter into S-R associations. Whereas, in the S-R condition, each rewarded action was reliably preceded by a particular arbitrary cue, rewarded actions in the R-O condition were reliably preceded by the emptying of a particular beaker. However, critically, it was only in the R-O condition that alternative actions could be associated with distinct outcome states. Consequently, we expected performance in this condition to be goal-directed.

Experimental procedure

We scanned participants as they acquired and performed the instrumental task, as well as during a subsequent devaluation test phase. Each subject participated in both the R-O and S-R condition, with conditions being run in separate, immediately consecutive, sessions (order counterbalanced across subjects) and with a novel set of four instrumental actions being used in each condition. Each condition included a response pre-training phase, learning phase, devaluation phase, and final test phase as described below. The response-training phase of the first condition was conducted outside the scanner (in a separate testing room), before participants were transferred to the scanner in which they remained throughout all subsequent stages of the experiment. Before being transferred to the scanner, participants were presented with a cover story describing the beaker system and task. They were told that they would be in one of two possible conditions (one in which each instrumental action refilled a particular beaker regardless of which cue was presented, and another in which the identity of the cue determined which of the four actions was required to refill an emptied beaker, regardless of the identity of that beaker) and that part of their task was to determine which of the two conditions they were in. The entire experiment lasted for approximately 2 h, with 1.5 h being spent in the scanner, and with approximately 60 min of active scanning during the learning phases and devaluation tests in each condition.

© 2015 Federation of European Neuroscience Societies and John Wiley & Sons Ltd

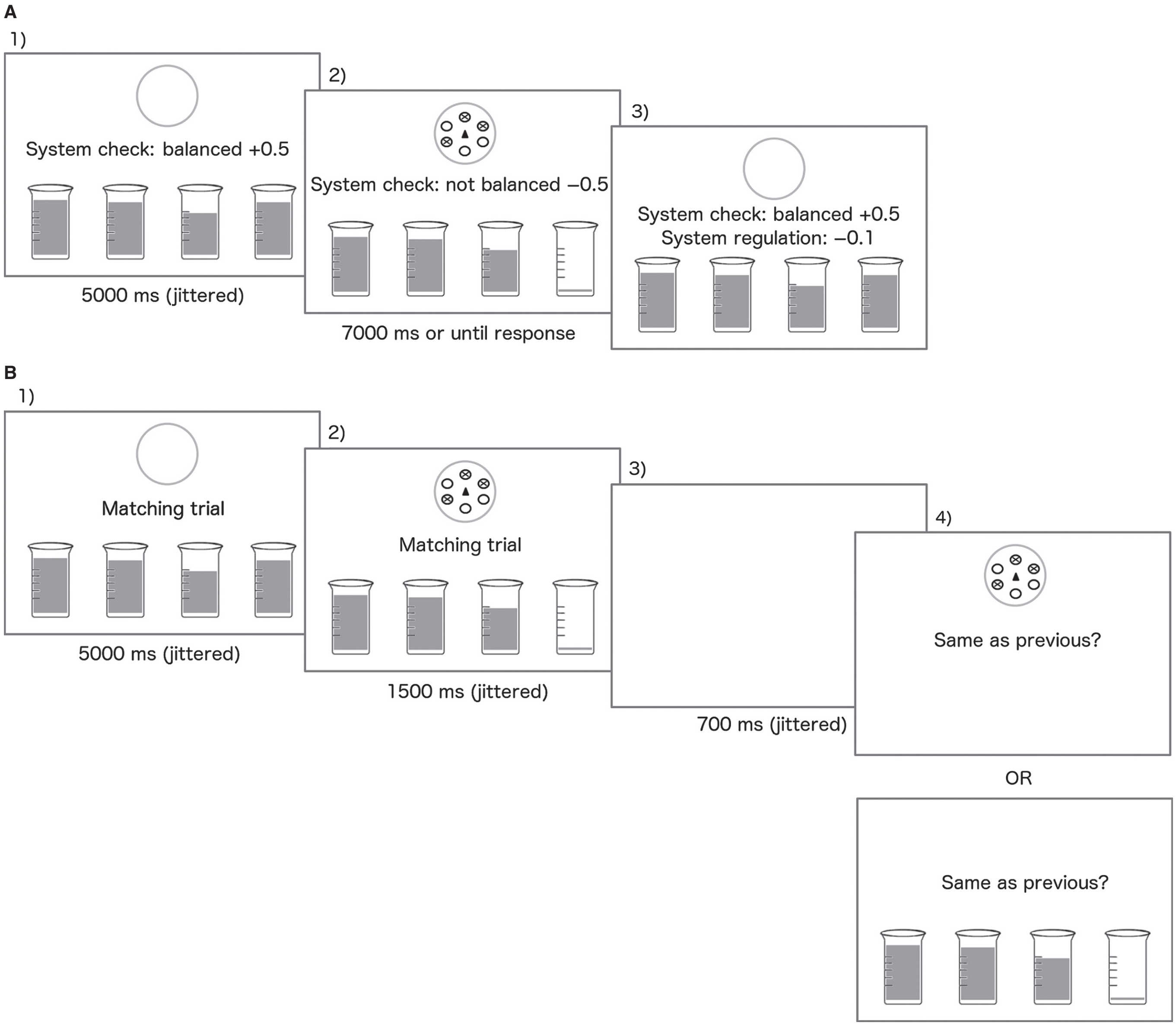

FIG. 1. Illustrations of instrumental learning and matching control tasks. (A) Instrumental learning. Participants are required to maintain the balance of a system of fluid-filled beakers using a set of four instrumental actions (each a three-press sequence, see Materials and methods). During the inter-trial interval, the liquid in the beakers continually fluctuates but remains high, and 'balance checks', occurring at brief random intervals, yield points for system balance (1). At the trial onset, one of four abstract cues appears, the liquid in one of the beakers drops to its bottom, and balance checks begin to indicate a loss of points due to system imbalance (2). Points are continually lost until the participant successfully refills the emptied beaker using one of the four actions. Following completion of the correct action (3), the abstract cue disappears, the beaker is refilled, a small fee is charged for regulating the system, and balance checks again yield points for system balance. If the correct action is not performed within 7 s, the beaker is automatically refilled, in which case there is no charge for system regulation. (B) Matching task. The inter-trial interval (1) and trial onset (2) were as in the instrumental task but were followed, at 1500 ms after trial onset, by a blank screen for 700 ms (3). The subsequent, final screen (4) showed a matching/non-matching stimulus together with a query about the match. In the S-R condition, the final screen always showed one of the abstract cues (top); conversely, in the R-O condition, the stimulus to be matched was always a set of beakers (bottom).

Response pre-training

Prior to the instrumental learning phases, participants received pre-training on the four instrumental actions (each being a three-press sequence). During this training, key-press sequences were illustrated by a white dot moving across three gray squares, horizontally aligned at the center of the screen. Initially, participants viewed and then immediately attempted to replicate each sequence, with feedback (i.e. correct/incorrect) given on each trial. After a total of five correct replications of each response sequence, they proceeded to a retrieval phase, in which they had to generate each unique sequence at least five times without any prompts, again with feedback given at the completion of each sequence. Participants were allowed to repeat these two phases as many times as they wanted to, knowing that they would have to use the actions to earn monetary reward in a subsequent phase.

Instrumental learning phase

The instrumental task was as illustrated in Fig. 1A. Note that, in addition to the increase or decrease in monetary points based on

© 2015 Federation of European Neuroscience Societies and John Wiley & Sons Ltd

system balance, there was a small cost for regulating the system. This response cost was included to ensure that, during test, participants would not respond simply based on any reinforcement intrinsic to executing the correct response. The response cost message (screen 3 in Fig. 1A) also served to inform participants that their action had successfully regulated the system, rather than the system having self-regulated (in which case no response cost was charged) due to failure to perform the correct action within 7 s of trial onset. Participants were allowed to perform as many key-presses as they wanted during system imbalance (i.e. during the 7 s following trial onset), with the correct sequence of three consecutive key-presses immediately restoring system balance and terminating the trial. There was no constraint on the temporal spacing between key-presses, as long as all three presses were performed consecutively within the trial window.

Critically, the stimulus materials presented in Fig. 1A were identical across the two conditions; our manipulation consisted entirely of differences in the contingencies between cues, actions and beaker outcomes. To rule out visual processes involved in selectively attending to the abstract cues vs. the beakers as a source of any imaging effects, a matching task was block-interleaved with the instrumental task during instrumental learning (see Fig. 1B). Briefly, in matching blocks, the inter-trial intervals and trial onsets were exactly as in the system balance task, except that the words 'Matching trial' were displayed center screen. Without this indication, matching trials would be identical to instrumental trials during the relevant trial period, and thus could not have served as controls. Following the appearance of the abstract cue and emptying of the relevant beaker, a white masking screen was displayed, followed by a depiction of either an abstract cue (S-R condition) or a set of beakers (R-O condition), together with a query about whether the currently shown cue/beaker set matched that on the previous screen. In each condition, the instrumental learning phase consisted of four blocks of trials, with each block being further divided into one sub-block of 24 instrumental trials, followed by a sub-block of eight matching trials, and with the order of trials randomized within each sub-block.

At the end of the instrumental learning phase, participants were asked whether they felt confident that they had learned how to regulate the system or whether they wanted to receive an additional set of 20 training trials (five with each action). Five participants (two in the R-O and three in the S-R condition) requested and received additional training. No scanning was conducted during additional training, nor were the added trials included in assessments of accuracy and response times during acquisition.

Devaluation phase

Following instrumental learning, participants were instructed that the system had changed such that one of the beakers was no longer relevant for system balance, which would be maintained, and continued to yield points even when the liquid in this beaker dropped below threshold. They then observed as the system regulated itself (i.e. no actions were performed) across 16 trials (four with each beaker) in order to discover the identity of the devalued beaker. Participants were told that they would not lose or gain any of the displayed points during this phase. In the imaging experiment, two participants failed to correctly identify the devalued beaker after this devaluation procedure, which was therefore repeated once for these participants. All other participants successfully identified the devalued beaker at the end of the devaluation procedure.

Test phase

Having correctly identified the devalued beaker, participants were again given the opportunity to regulate the system for personal monetary reward. During this phase, all text messages indicating gains or losses were covered up, in order to prevent additional learning (i.e. simulating extinction). Participants were instructed that, in spite of these gray strips, they should assume that all was exactly as they had learned before, i.e. they would still lose points whenever the system was not balanced, there was still a cost for regulating the system, and the previously identified irrelevant beaker was still irrelevant for system balance. Importantly, because there was a small charge for each instrumental regulation of the system, refilling the now irrelevant beaker resulted in a net monetary loss. The test phase consisted of a single instrumental block with 44 randomly ordered trials, i.e. 11 trials with each beaker, including the devalued (i.e. irrelevant) one. We operationally defined devaluation insensitivity, our assay of habitual performance, as the proportion of the 11 devalued trials on which participants initiated a response to refill the beaker. Thus, if a participant initiated a response on two of those 11 trials, their devaluation insensitivity score would be 0.18.

Pay-off structure

Throughout the instrumental task, system balance checks occurred on average every 3 s. During the inter-trial interval (mean 5 s), the system was always balanced, with each check yielding a reward of 0.5 points. At each trial onset, the system became imbalanced, with each balance check resulting in a loss of 0.5 points, and would remain so for 7 s unless regulated by the participant. Thus, if no action were taken to balance the system, each trial would entail an average loss of 1.17 points. If the system was balanced immediately, no points were lost due to system imbalance, but there was a response cost of 0.1 points (this cost was only applied to correct responses resulting in system regulation). Thus, the average gain of responding to balance the system was 1.07 points. After beaker devaluation, during the test phase, this was still the case for regulatory responses refilling non-devalued beakers; however, responding to refill the devalued beaker now resulted in a net loss of 0.1 points (the response cost).

Imaging acquisition and analyses

A 3 Tesla scanner (MAGNETOM Trio, Siemens) was used to acquire structural T1-weighted images and T2*-weighted echoplanar images (repetition time, 2.65 s; echo time, 30 ms; flip angle, 90°; 45 transverse slices; matrix, 64 × 64; field of view, 192 mm; thickness, 3 mm; slice gap, 0 mm) with blood oxygenation level-dependent contrast. To recover signal loss from dropout in the medial orbitofrontal cortex (O'Doherty *et al.*, 2002), each horizontal section was acquired at 30° to the anterior commissure–posterior commissure axis. Image processing and statistical analyses were performed using statistical parametric mapping (SPM)8 (http://www.fil.ion.ucl.ac.uk/spm). The first four volumes of images were discarded to avoid T1 equilibrium effects. All remaining volumes were corrected for differences in slice acquisition, realigned to the first volume, spatially normalized to the Montreal Neurological Institute echoplanar imaging template, and spatially smoothed with a Gaussian kernel (8 mm, full width at half-maximum). We used a high-pass filter with a cutoff of 128 s.

As previously noted, instrumental conditions were run in separate consecutive sessions, with the acquisition phase of each condition

© 2015 Federation of European Neuroscience Societies and John Wiley & Sons Ltd

consisting of interleaved instrumental and matching block, and with active scanning occurring only during the acquisition and test phase of each condition (i.e. the scanner was off during response pre-training and devaluation phases), for a total of four active scanning periods per subject. For each subject, we constructed a design matrix that included the acquisition and test phases of both experimental conditions. For each of the two instrumental acquisition phases (i.e. R-O and S-R), stick functions modeled the trial onsets in each block of instrumental learning and in matching blocks. In addition, three regressors, respectively modeling the onsets of error trials, the times of all individual key-presses and the times of all point displays (i.e. gains and losses), were entered as regressors of no interest, together with six regressors accounting for the residual effects of head motion. For each of the two test phases, two stick functions respectively modeled non-devalued and devalued trials. Again, for each instrumental condition, additional regressors modeling the onsets of error trials and the times of all key-presses were added together with six regressors accounting for the residual effects of head motion as regressors of no interest. All regressors were convolved with a canonical hemodynamic response function.

Group-level random-effects statistics were generated by entering contrasts of parameter estimates for the different regressors into between-subjects analyses of variance (ANOVAS). Specifically, to delineate neural substrates engaged during early acquisition vs. during expression, contrasts of parameter estimates for the first two blocks of instrumental learning, and for the devaluation test phase, were entered for each instrumental condition into a 2 × 2 (condition by experimental phase) ANOVA. Contrasts of parameter estimates for all learning blocks and for matching blocks were entered into a separate 2 × 2 (stimulus by task) ANOVA to compare differences between instrumental conditions during learning with those between matching control conditions. An analogous ANOVA was performed using contrasts of parameter estimates for devaluation test trials and matching trials. Finally, a 4 × 2 (acquisition block by instrumental condition) ANOVA assessed training-related changes in neural activity, with contrasts of parameter estimates for each of the four blocks of instrumental learning, for each instrumental condition. Following estimation of the second-level model, F-tests were specified by adding linear weights to each block of learning (e.g. [−1.5 −0.5 0.5 1.5] for increases across blocks and [1.5 0.5 −0.5 −1.5] for decreases) in each instrumental condition.

To assess whether neural discrimination between instrumental conditions correlated with the degree of devaluation insensitivity, a simple *t*-test was performed on first-level interaction contrasts [S-R > R-O × Instrumental > Matching] of parameter estimates from the learning phase, with the degree of devaluation insensitivity entered as a covariate. An analogous *t*-test using interaction contrasts of parameter estimates from the devaluation test phase was also performed, and exclusive functional masks were used to assess the specificity of neural effects, such that all voxels that reached significance at a threshold of $P < 0.1$ when assessing correlates of devaluation insensitivity during the test phase were removed from the effects observed during learning, and vice versa. Exclusive functional masks were also used to assess the directions of simple effects underlying the Instrumental Condition by Experimental Phase interaction. Finally, we used exclusive functional masks to selectively assess training-related increases vs. decreases in neural activity, such that, for example, when testing for increases in neural activity across blocks of training in the R-O condition, voxels were removed that reached significance at $P < 0.1$ for the same test in the S-R condition.

Small volume corrections (SVCs) were performed on several *a-priori* regions of interest using a 10 mm sphere, with center coordinates obtained from highly relevant studies. Specifically, in an analogous study on observational learning (Liljeholm *et al.*, 2012), we found selective recruitment of the extrastriate cortex (±45, −72, 9), the tail of the caudate nucleus/thalamus (±21, −30, 9) and the lingual gyrus (±12, −69, 0) by the S-R condition during acquisition. Conversely, areas selectively recruited by the R-O condition in that same observational learning study, and also identified in our other work on goal-directed performance (Liljeholm *et al.*, 2011), include the inferior parietal lobule (IPL) (±51, −52, 33) and anterior caudate nucleus (±16, 8, 19). Finally, several human neuroimaging studies have implicated a posterior region of the putamen (±33, −24, 0) in habit formation (Tricomi *et al.*, 2009; de Wit *et al.*, 2012; Wunderlich *et al.*, 2012; Lee *et al.*, 2014), and the ventromedial prefrontal cortex (±4, 55, −7) in goal-directed processes (Hampton *et al.*, 2006; Tanaka *et al.*, 2008; Liljeholm *et al.*, 2011). It should be noted that, although specified *a priori* based on a closely related literature, effects in several of these regions survived whole-brain cluster-size thresholding (CST) correction, as well as SVC, in the current study.

As noted, the putamen, considered a human homolog of the rodent dorsolateral striatum based on its afferent and efferent projections (Carelli & West, 1991; Draganski *et al.*, 2008), has been implicated in habitual performance in several human imaging studies (Tricomi *et al.*, 2009; de Wit *et al.*, 2012; Wunderlich *et al.*, 2012; Lee *et al.*, 2014). In contrast, the role of the infralimbic cortex in instrumental performance has been exclusively demonstrated in rodents (Killcross & Coutureau, 2003; Haddon & Killcross, 2011; Smith *et al.*, 2012; Smith & Graybiel, 2013). Consequently, and based on anatomical and functional evidence for a homology between the rodent infralimbic cortex and the human subgenual cortex (Ongur & Price, 2000; Ongur *et al.*, 2003; Drevets *et al.*, 2008), predictions regarding this area were tested using an anatomical mask of the subgenual cortex (Brodmann area 25), defined using the Wake Forest University Pickatlas (Maldjian *et al.*, 2003). All other effects were reported at $P < 0.05$, using CST of SPM t-maps to adjust for multiple comparisons (Forman *et al.*, 1995). AlphaSim (Ward, 2000), a Monte Carlo simulation, was used to determine cluster size and significance. Using an individual voxel probability threshold of $P = 0.005$ indicated that using a minimum cluster size of 111 Montreal Neurological Institute transformed voxels resulted in an overall significance of $P < 0.05$.

To eliminate non-independence bias for plots of contrast values, a leave-one-subject-out (Esterman *et al.*, 2010) approach was used, in which 19 general linear models were run with one subject left out in each, and with each general linear model defining the voxel cluster for the left-out subject. Using rfxplot (Glascher, 2009), mean contrast values were extracted from spheres (10 mm) centered on these leave-one-subject-out peaks (identified within regions of interest for SVCs) and were averaged across subjects to plot overall effect sizes.

Results

Behavioral results

The results from the test phase indicated that our manipulation did indeed produce differences in devaluation sensitivity. Having correctly identified the devalued beaker, participants initiated a response on a significantly greater proportion of devalued trials in the S-R condition (mean proportion 0.43, SEM 0.09) than in the R-O condition (mean proportion 0.19, SEM 0.09) [t_{18} = 2.3, $P < 0.031$]. Indeed, whereas only five of 19 participants initiated a response on any devalued trials in the R-O condition, 15 of 19 participants initiated a response on at least one devalued trial in the S-R condition ($\chi^2 = 10.56$, $P < 0.005$).

Note that devaluation insensitivity was defined such that a response to obtain the devalued outcome (i.e. to refill the devalued beaker) was counted even if the entire three-press sequence was not completed on that trial. Indeed, in the S-R condition, when a participant responded to fill up the devalued beaker, they often did so without completing the entire three-press sequence, consistent with evidence from the rodent literature that the response most proximal to reward remains sensitive to devaluation long after more distal responses have become habitual (Killcross & Coutureau, 2003). In contrast, those few participants that responded on devalued trials in the R-O condition did so on most or all devalued trials and completed all three key-presses whenever initiating a response, suggesting a more deliberate decision to respond.

Importantly, during the test phase, whereas in the R-O condition participants could determine both the accuracy and value of a particular action solely by identifying the emptied beaker, in the S-R condition, determining the value and accuracy of a given response required identification of both the abstract cue and the emptied beaker. It is possible therefore that the differences in devaluation performance were due to this additional aspect of the S-R condition. The overall pattern of behavioral results, however, strongly suggests that this was not the case, and generally rules out task difficulty as the source of our effects. First, there were no significant differences between conditions in the percent of incorrect responses, during either acquisition [R-O: mean 14%, SEM 3; S-R: mean 11%, SEM 2; $t_{18} = 1.1$, $P = 0.30$] or test performance [R-O: mean 4%, SEM 1; S-R: mean 4%, SEM 2; $t_{18} = 0$, $P = 1.0$]. Likewise, response times did not differ between conditions during either acquisition [R-O: mean 1482 ms, SEM 62; S-R: mean 1393 ms, SEM 78; $t_{18} = 1.95$, $P = 0.21$] or test performance [R-O: mean 1258 ms, SEM 54; S-R: mean 1361 ms, SEM 68; $t_{18} = -1.52$, $P = 0.15$]. The fact that there were no significant differences between conditions in either accuracy or reaction times during test makes it highly unlikely that differences in devaluation sensitivity reflected differences in task difficulty. Perhaps most pertinently, differences in devaluation insensitivity were not correlated with differences in accuracy ($P = 0.5$) or reaction times ($P = 0.8$), nor did individual differences in accuracy or reaction times predict neural effects in any of the areas identified by our imaging analyses.

To assess the influence of counterbalancing order, we performed order-by-condition analyses of variance on response times and accuracy scores during the learning and test phases, as well as on the devaluation insensitivity scores. There was no significant interaction between order and condition for devaluation insensitivity ($F_{1,17} = 0.55$, $P = 0.47$), or for either response times ($F_{1,17} = 0.89$, $P = 0.36$) or accuracy scores ($F_{1,17} = 0.07$, $P = 0.79$) during the test phase. In contrast, during training, there was an anticipated interaction for both response times ($F_{1,17} = 5.96$, $P < 0.03$) and accuracy scores ($F_{1,17} = 23.01$, $P < 0.001$), such that response times were longer and the percent incorrect was greater for whichever condition came first. Thus, whereas the difference in response times between instrumental conditions differed across the two orders (mean difference for R-O first 0.19; mean difference for S-R first −0.32), there was no difference between instrumental conditions when compared for a given order (i.e. for R-O vs. S-R with both first, $P = 0.84$ and for R-O vs. S-R with both second, $P = 0.71$). A similar pattern was observed for percent incorrect scores (mean difference for R-O first 0.08; mean difference for S-R first −0.01; significance test for R-O vs. S-R with both first, $P = 0.26$ and for R-O vs. S-R with both second, $P = 0.90$). Consequently, we can rule out counterbalancing order as a source of our behavioral and imaging effects of interest.

There were no differences between instrumental conditions in the number of requested repetitions of response pre-training phases [R-O: mean 1.98, SEM 0.29; S-R: mean 1.75, SEM 0.18; $t_{18} = 0.72$, $P = 0.48$], or in the total number of key-presses executed during instrumental learning [R-O: mean 470.63, SEM 32.14; S-R: mean 503.74, SEM 31.63; $t_{18} = -0.85$, $P = 0.40$] or on non-devalued trials during test [R-O: mean 121.84, SEM 4.57; S-R mean: 125.53, SEM 4.53; $t_{18} = -1.34$, $P = 0.57$]. Finally, there was no difference between instrumental conditions in the total number of points earned during instrumental learning [R-O: mean 33.6, SEM 3.61; S-R: mean 33.4, SEM 3.33; $t_{18} = 0.071$, $P = 0.94$]. This allowed us to rule out the number of motor responses as well as total earnings as sources of the difference in devaluation performance.

Neuroimaging results

Our primary objective was to investigate whether neural discrimination between the two instrumental conditions differed across the acquisition and expression of behavioral control, particularly with respect to blood oxygenation level-dependent activity predictive of individual differences in devaluation insensitivity (see Table 1). To rule out the possibility that differences between our instrumental conditions were due to differences in the processing of relevant visual features (i.e. of abstract cues in the S-R condition and of the set of beakers in the R-O condition), additional analyses formally contrasted such differences with those emerging between matching control conditions (see Table 2). Unless otherwise noted, all effects reported below survived the subtraction of matching control conditions. All of the results described below survived correction for multiple comparisons at $P < 0.05$ using either whole-brain CST or SVCs based on coordinates from relevant previous studies. Notably, many of the areas specified *a priori* as targets of SVC, including the lingual gyrus, extrastriate cortex and IPL, also survived correction using whole-brain CST, as indicated in the fourth columns of Tables 1 and 2.

Discrimination between instrumental conditions during acquisition vs. performance

To delineate the neural substrates engaged during early acquisition vs. during expression of operant behavior, we entered the imaging data from the first two blocks of instrumental learning together with that from the subsequent test phase, for each instrumental condition, into a 2 × 2 (condition by experimental phase) ANOVA (see Table 1). Although processes supporting acquisition should certainly dominate during the first two blocks of instrumental learning, we cannot completely rule out the possibility that some element of expression was also present during this phase. Inclusion of only the very earliest learning trials would not eliminate this possibility, but would introduce dramatically different error levels and severely reduce the statistical power. We note, however, that the presence of processes related to expression during the early stages of instrumental learning should reduce, rather than enhance, discrimination between the levels of the ‘experimental phase’ variable. Thus, in areas where we did observe a significant difference in discriminatory activity across the different phases of the experiment, such differences were probably attributable to the differential effects of acquisition vs. expression.

© 2015 Federation of European Neuroscience Societies and John Wiley & Sons Ltd

TABLE 1. Imaging results from a 2 × 2 ANOVA contrasting instrumental conditions and training phases

Test	Area	x, y, z	T at peak level	Correction*	Cluster size at $P < 0.005$*
Main effects					
S-R > R-O	Inferior occipital	36, −85, −8	6.33	CST	859
	Middle occipital	36, −85, 7	6.82		
	Superior occipital	27, −67, 28	5.16		
	SPL	18, −61, 52	4.22		
R-O > S-R	DMPFC	18, 56, 25	3.62	CST	843
	VMPFC	3, 41, 1	3.25	CST	
	IPL	60, −55, 28	3.99	CST	187
	Posterior cing.	−3, −34, 43	3.13	CST	190
	Insula	39, 14, −41	3.89	CST	157
Acq. > Perf.	VMPFC	−3, 62, 4	6.12	CST	2181
	DMPFC	−3, 47, 37	5.65		
	VS	−6, 5, −8	4.90		
	aCN	9, 20, 7	3.53		
	VLPFC	−45, 32, −11	5.46	CST	354
	IPL	−51, −64, 34	4.76	CST	200
Perf. > Acq.	Cerebellum	27, −61, −26	9.17	CST	9773
	SMA	0, 8, 46	5.15		
	Insula/putamen	−33, 8, 4	5.69		
	Post-central S.	−42, −37, 58	6.33		
	Pre-central S.	−30, −4, 52	5.97		
	Calcarine	24, −58, 10	4.94		
	SPL	−15, −64, 49	5.31		
Interactions					
S-R > R-O (Perf.>Acq.)	Occipital	33, −79, 19	5.10	CST	859
	SPL	24, −52, 46	3.58		
R-O > S-R (Perf.>Acq.)	DMPFC	−9, −41, 40	4.26	CST	168

*The absence of an entry in the ‘correction’ and ‘cluster size’ columns indicates that the relevant cluster is continuous (at the $P < 0.005$ threshold) with that listed above it. Acq., Acquisition; Perf., Performance; SPL, superior parietal lobule; DMPFC, dorsomedial prefrontal cortex; aCN, anterior caudate nucleus; cing., cingulate; VLPFC, ventrolateral prefrontal cortex; SMA, supplementary motor area; S., post-/pre-central sulcus.

TABLE 2. Imaging results from 2 × 2 ANOVAS contrasting instrumental and matching tasks, and from tests of neural correlates of devaluation insensitivity

Contrast	Area	x, y, z	T at peak level	Correction*	Cluster size at $P < 0.005$*
Learning phase					
S-R > R-O (instr. > ctrl)	EBA	39, −79, 10	4.09	SVC	45
R-O > S-R (instr. > ctrl)	IPL	57, −58, 31	4.41	CST	275
	Insula	33, 14, 4	3.57	CST	421
	Precuneus	6, −70, 37	4.28	CST	395
	Post. cing.	9, −31, 43	3.36		
	aCN	12, 2, 16	3.40	SVC	18
Test phase					
S-R > R-O (instr. > ctrl)	Occipital	30, −79, 19	5.30	CST	1153
	SPL	24, −62, 49	4.47		
R-O > S-R (instr. > ctrl)	IPL	57, −46, 37	3.09	SVC	18
	Insula	42, −1, −5	4.06	CST	383
	DMPFC	−6, 41, 40	3.87	CST	382
Matching effects					
S-R control > R-O control	Lateral LG	24, −91, −8	5.20	CST	198
R-O control > S-R control	Medial LG	−9, −76, −8	5.15	CST	1541
	Cuneus	−9, −91, 19	5.01		
Correlates of devaluation insensitivity					
Positive correlation	CRBL/LG	−6, −52, −8	5.06	CST	125
	tCN/th	−12, −28, 10	4.19	SVC	21
	Subgenual	−15, 23, −14	5.24	CST	224
	VS	−9, 5, −14	8.47		
Negative correlation	IPL	45, −43, 34	4.62	SVC	54

*The absence of an entry in the ‘correction’ and ‘cluster size’ columns indicates that the relevant cluster is continuous (at the $P < 0.005$ threshold) with that listed above it. Instr., instrumental; ctrl, control; Post. cing., posterior cingulate; aCN, anterior caudate nucleus; SPL, superior parietal lobule; DMPFC, dorsomedial prefrontal cortex; tCN/th, tail of the caudate nucleus/thalamus; CRBL, cerebellum; LG, lingual gyrus.

Stimulus–Response-related activity

Activity that was greater in the S-R than the R-O condition (Fig. 2) emerged in the middle and superior occipital cortex, and in the superior parietal lobule. Interaction tests and (exclusively masked) tests of simple effects revealed that activity in the superior parietal lobule and superior occipital cortex was significantly greater in the

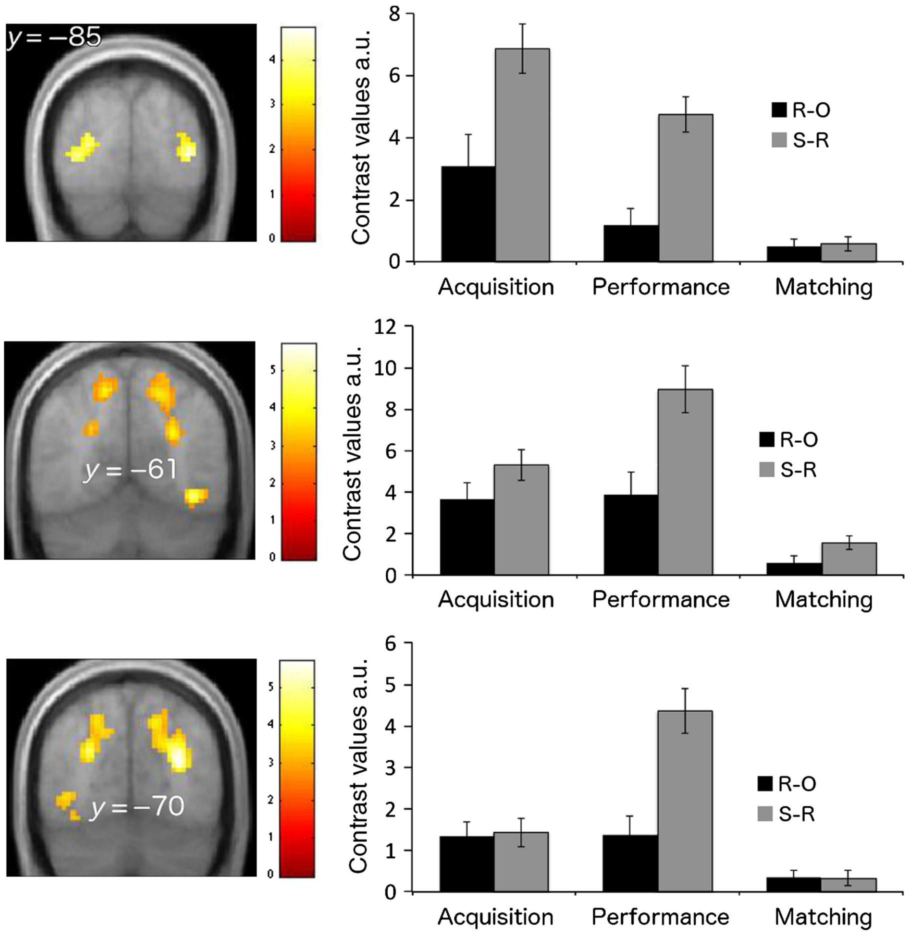

FIG. 2. Areas selectively involved in the S-R condition. Statistical maps show main effects of the S-R > R-O contrast, in the middle occipital cortex (top row), and an interaction effect {i.e. [S-R > R-O (performance > acquisition)]} in the superior occipital cortex, extending into the superior parietal lobule (middle and bottom rows). Bar plots show contrast values estimated at leave-one-subject-out coordinates for each instrumental condition, during both acquisition and performance, and for matching controls. Error bars = SEM. a.u., arbitrary units.

S-R than the R-O condition only during test performance, but not during early acquisition (Table 1). Moreover, during early acquisition, only a small area of the middle occipital cortex, the extrastriate body area (EBA; SVC), survived subtraction of corresponding matching controls (Table 2). In contrast, S-R selective activity throughout the occipital cortex, including the EBA, survived subtraction of matching controls during the test phase.

Response–Outcome-related activity

Activity significantly greater in the R-O condition than the S-R condition (Fig. 3) was observed throughout a fronto-parietal network, including the dorsal and ventral medial prefrontal cortex (VMPFC), IPLs, posterior cingulate and bilateral insula. Interaction tests and (exclusively masked) tests of simple effects revealed that, in the dorsomedial prefrontal cortex, discrimination between instrumental conditions occurred during test performance only (Table 1). Analyses assessing differences between conditions relative to matching controls also revealed R-O selective effects in the dorsal anterior caudate, but did not yield significant effects in the VMFPC (Table 2).

Discrimination between conditions across blocks of acquisition

To further assess differential activity related to acquisition processes, we tested for differences between conditions in training-dependent changes in neural activity, by adding linear weights to blocks of instrumental learning. An interaction test assessing neural activity that increased in the R-O condition and decreased in the S-R condition yielded significant effects throughout the right putamen and globus pallidus (CST, 33, −7, −5) and in the right IPL (CST, 63, −31, 37). The reverse interaction contrast, assessing neural activity that decreased in the R-O condition and increased in the S-R condition, did not reveal any significant effects. Specific assessments of increases vs. decreases revealed decreases across training blocks in the S-R condition, but not in the (exclusively masked) R-O condition, throughout the right putamen and globus pallidus (Fig. 4A). In contrast, significant increases in activity across blocks in the R-O condition, but not in the (exclusively masked) S-R condition, emerged in the right IPL (Fig. 4B).

Neural correlates of devaluation insensitivity

To relate the neuroimaging data to our behavioral effects, we tested whether neural discrimination between instrumental conditions correlated with differences between conditions in the degree of devaluation insensitivity. This was indeed the case. Participants with stronger activation of the tail of the caudate nucleus/thalamus and of the cerebellum, extending into the lingual gyrus, in the S-R relative to the R-O condition during the first two blocks of instrumental learning responded on a greater proportion of devalued trials in the

© 2015 Federation of European Neuroscience Societies and John Wiley & Sons Ltd

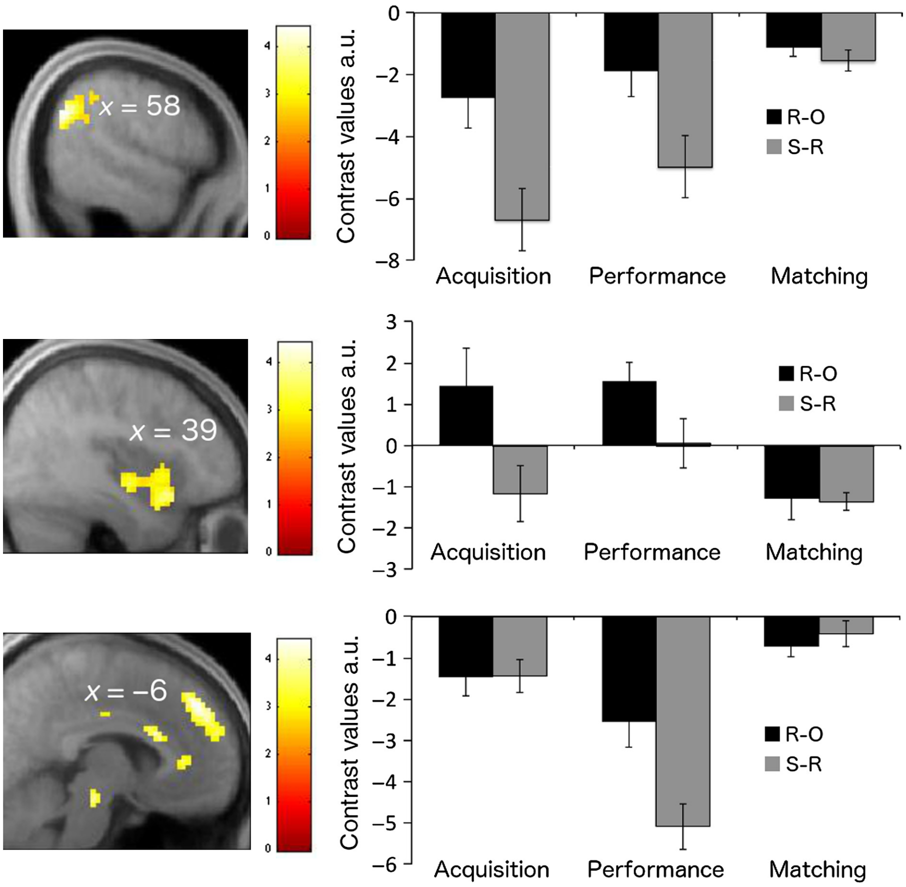

FIG. 3. Areas selectively involved in the R-O condition. Statistical maps show main effects of the R-O > S-R contrast in the posterior ventral IPL (top) and insula (middle), as well as an interaction effect {i.e. [R-O > S-R (performance > acquisition)]} in the dorsomedial prefrontal cortex (bottom). Bar plots show contrast values estimated at leave-one-subject-out coordinates for each instrumental condition, during both acquisition and performance, and for matching controls. Error bars = SEM. a.u., arbitrary units.

S-R condition relative to the R-O condition during the subsequent test phase (Fig. 5A). These effects all survived exclusive masking by an identical contrast applied, with a threshold of 0.1, to the imaging data from the test phase, suggesting that, in these areas, discriminatory activity during early acquisition, but not during test performance, predicted differences in devaluation insensitivity.

Conversely, during the test phase, differences in neural activity between the S-R and R-O conditions in the subgenual cortex (region of interest) and ventral striatum (VS) were positively correlated with the difference between conditions in devaluation insensitivity (Fig. 5B), i.e. greater subgenual and VS activity in the S-R relative to the R-O condition predicted greater devaluation insensitivity in the S-R than R-O condition. We also found a negative correlation with test performance in the right IPL (SVC), such that lesser IPL activity in the S-R relative to the R-O condition predicted greater devaluation insensitivity in the S-R than R-O condition. Again, the specificity of these results was confirmed using exclusive masking by an identical contrast, at a threshold of 0.1, applied to the imaging data from the training phase.

Discussion

In this study we explored the neural substrates of goal-directed and habitual action selection, with a focus on how the recruitment of relevant brain areas might differ across acquisition and implementation of behavioral control. We scanned human participants with functional magnetic resonance imaging as they learned and performed a novel task designed to encourage either goal-directed encoding of the specific outcomes of instrumental responses (R-O condition), or a habitual mapping of responses to antecedent cues (S-R condition). In a subsequent test phase, participants were more likely to respond for a devalued outcome, indicative of habits, in the S-R condition. We found that neural activity in striatal and cortical areas (i) discriminated between our two instrumental conditions, (ii) predicted individual differences in devaluation insensitivity and (iii) did so differentially across acquisition and test performance.

Our results identified several areas in which the degree of neural discrimination between instrumental conditions predicted differences in devaluation insensitivity. Specifically, S-R selective activity in the tail of the caudate nucleus and cerebellum during learning, but not during test performance, predicted greater devaluation insensitivity in the S-R, relative to the R-O, condition. The cerebellum has been strongly implicated in response automatization (Doyon *et al.*, 1998, 2002; Lang & Bastian, 2002; Balsters & Ramnani, 2011), a resistance to dual-task interference postulated to be closely related to, and sometimes treated as synonymous with, habitual performance. Indeed, in the rodent literature, the cerebellum has been directly implicated in habit formation, such that cerebellar lesions abolish devaluation insensitivity in overtrained rats (Callu *et al.*, 2007). However, to our knowledge, no previous study has directly linked the cerebellum to outcome devaluation insensitivity in humans. As with the cerebellum, the tail of the caudate nucleus has been

© 2015 Federation of European Neuroscience Societies and John Wiley & Sons Ltd

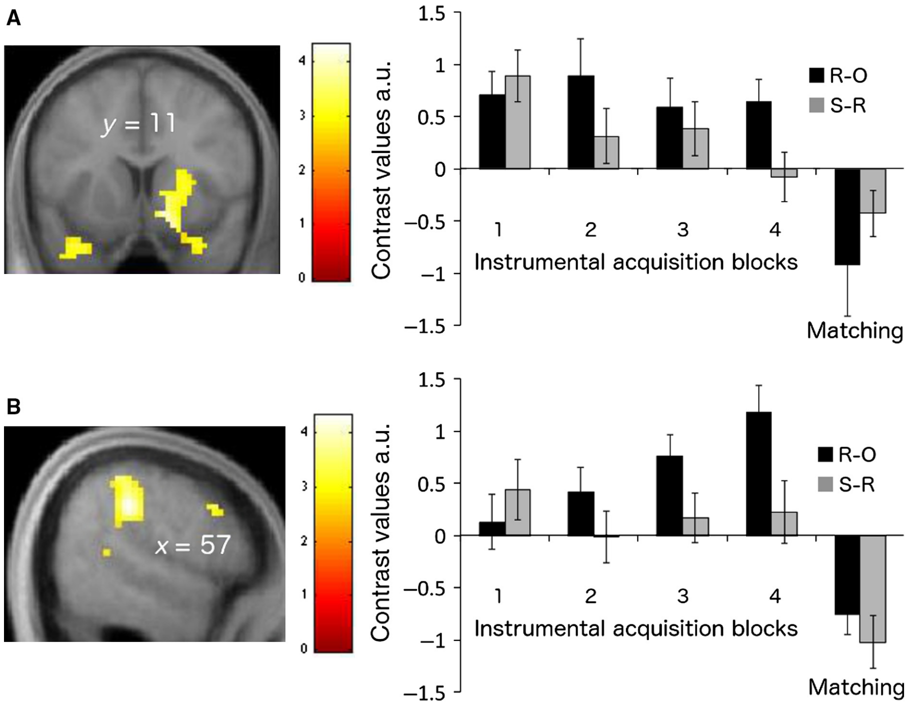

FIG. 4. Discrimination between conditions across blocks of acquisition. (A) Statistical map showing results from disjunction tests assessing decreases across blocks of acquisition in the S-R but not the R-O condition with effects emerging in the right putamen/globus pallidus. (B) Results from a disjunction test of increases across blocks in the R-O but not the S-R condition yielded effects in the IPL. Bar plots show contrast values estimated at leave-one-subject-out coordinates for each instrumental condition in each training block, as well as for matching control conditions. Error bars = SEM. a.u., arbitrary units.

implicated in skill learning (Poldrack & Gabrieli, 2001; Yamamoto *et al.*, 2013). Moreover, single neuron recordings in the monkey tail of the caudate nucleus have revealed that, relative to the body and head of the caudate, this area is specifically involved in encoding stable, non-flexible, values of visual cues (Kim & Hikosaka, 2013; Yamamoto *et al.*, 2013; Hikosaka *et al.*, 2014; Kim *et al.*, 2014), consistent with its role in devaluation insensitivity demonstrated here and elsewhere (Valentin *et al.*, 2007; Liljeholm *et al.*, 2012).

In the subgenual cingulate (Brodmann area 25), S-R selective activity during test performance, but not during learning, correlated with behavioral devaluation insensitivity. Based on its cytoarchitectonic subdivisions and connections, this area has been identified as homologous to the rodent infralimbic cortex. The rodent infralimbic and human subgenual areas both project heavily to the shell of the nucleus accumbens, and both are agranular, relatively poorly laminated areas located ventrally on the medial wall (Gabbott *et al.*, 1997; Ongur & Price, 2000; Ongur *et al.*, 2003). In rodents, the infralimbic cortex has been shown to make specialized contributions to executive control processes facilitating the deployment of habits (Coutureau & Killcross, 2003; Haddon & Killcross, 2011; Smith *et al.*, 2012; Smith & Graybiel, 2013). Consistent with such findings, our results suggest that a putative human homolog of this area does indeed play selective roles in the expression of habits. It should be noted, however, that the rodent infralimbic cortex is primarily known for its involvement in the extinction of Pavlovian responses (Milad & Quirk, 2002; Rhodes & Killcross, 2004, 2007; Santini *et al.*, 2008), whereas the human subgenual cortex has predominantly featured in studies on depression (Greicius *et al.*, 2007; Drevets *et al.*, 2008; Johansen-Berg *et al.*, 2008; Matthews *et al.*, 2009; Keedwell *et al.*, 2010). Future work is needed to reconcile these apparently divergent functions and their potential homology across species.

As with the subgenual cortex, activity in the VS during the test phase, but not during acquisition, predicted devaluation insensitivity. The VS has been frequently shown to support both the acquisition and expression of Pavlovian (stimulus–outcome) associations (Day *et al.*, 2007; Blaiss & Janak, 2009), and to mediate the general motivational influence of such associations on instrumental performance, a phenomenon referred to as Pavlovian-instrumental transfer (Corbit & Balleine, 2011). Of particular importance for interpreting the current findings is the fact that, in rodents, Pavlovian-instrumental transfer appears to selectively influence habitual, rather than goal-directed, responding, i.e. the greater the degree of insensitivity to outcome devaluation, the greater the general motivational influence of Pavlovian cues on instrumental performance (Holland, 2004; Balleine & Ostlund, 2007). We interpret the currently observed activity in the VS as reflecting a greater influence of Pavlovian cues on instrumental responding in the S-R than the R-O condition, and conjecture that this selective engagement of Pavlovian processes supported our behavioral effect.

Contrary to our predictions, we did not find a correlation between activity in the dorsal putamen during early learning and subsequent devaluation insensitivity. We did, however, find S-R selective decreases in right putamen activity across blocks of training, an effect that is consistent with a substantial literature demonstrating a decreased dependence on the putamen with extended (Ungerleider *et al.*, 2002; Poldrack *et al.*, 2005; Turner *et al.*, 2005; Ashby *et al.*, 2010), as well as intermediate (Brovelli *et al.*, 2011), levels of training. Taken together, these results suggest that the contributions of the putamen to automatic and habitual behavioral control may take place early in acquisition, with long-term storage and mediation of well-trained performance occurring elsewhere (Orban *et al.*, 2010). However, some neuroimaging studies have found increases in putamen activity with extended training (Floyer-Lea & Matthews, 2004;

© 2015 Federation of European Neuroscience Societies and John Wiley & Sons Ltd

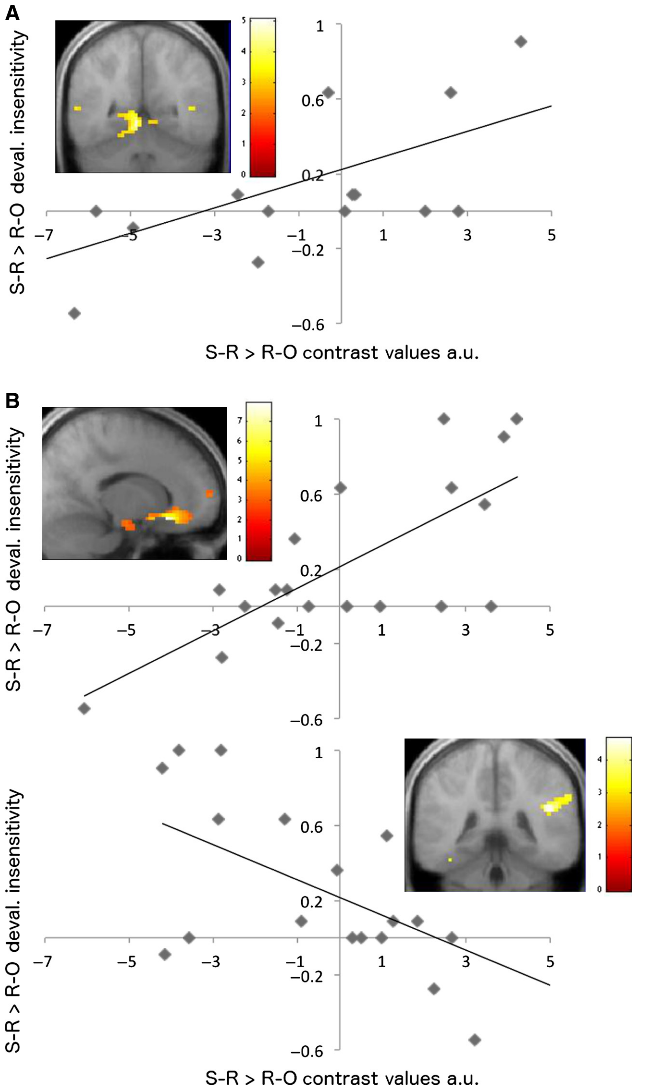

FIG. 5. Correlations between contrast values for the [S-R > R-O (instrumental > control)] contrast and behavioral differences between the S-R and R-O conditions in the proportion of devaluation insensitive responses. Contrast values are extracted from 8 mm spheres centered on the peak voxel in each area. (A) Correlation between contrast values from the first two blocks of instrumental learning (x-axis) and differences between instrumental conditions in the proportion of devaluation insensitive responses (y-axis), showing effects in the cerebellum. (B) Correlation between contrast values from the test phase (x-axis) and concurrent differences between instrumental conditions in the proportion of devaluation insensitive responses (y-axis), showing a positive correlation in the subgenual cortex (top) and a negative correlation in the right IPL (bottom).

Tricomi *et al.*, 2009; Wunderlich *et al.*, 2012). For instance, Tricomi *et al.* (2009) found an increase in activity with overtraining in a far posterior region of the right putamen, concomitant with the development of habitual performance. One key difference between the study of Tricomi *et al.* (2009) and the present one is that the S-R condition in the present study is optimized to generate the rapid acquisition and expression of habitual behavior, whereas in the study of Tricomi *et al.* (2009), behavioral expression of habits (and perhaps also acquisition) emerged much more slowly, becoming evident in behavior only after several days of training. Thus, if the involvement of the putamen in S-R learning dissipates after a period of stable habitual performance, Tricomi *et al.* (2009) may not have sampled behavior beyond that stable period, whereas our accelerated habitual learning paradigm allowed us to do so. It is also worth noting, that, whereas Tricomi *et al.* (2009) reported effects in a small area in the very far posterior putamen, our effects extend throughout the right putamen and globus pallidus.

One feature of the present results is that areas in which discriminatory neural activity was correlated with between-subject variation in devaluation insensitivity did not also show differential main effects in a comparison of S-R and R-O conditions. This may be due in part to the fact that our efforts to differentially encourage S-R and R-O learning resulted only in relative differences in devaluation insensitivity, with the majority of participants showing some degree of sensitivity even in the S-R condition. Thus, neural processes directly related to the degree of habitual responding might not be discernable in analyses that categorically compare conditions. However, not all aspects of encoding S-R associations necessarily contribute to the dominance of such associations over performance. For example, the S-R selective decreases in neural activity across blocks of acquisition, found throughout the right putamen and globus pallidus, may reflect a gradual disengagement of processes that provide critical support during the early development of S-R associations (such as the maintenance of relevant representations, the encoding of discrepancies between attempted and accurate mappings, or the retrieval of response alternatives) but that are supplanted during subsequent performance by processes mediating behavioral control.

We also found selective recruitment by the S-R condition of a lateral middle occipital area that overlaps closely with what has been termed the ‘EBA’ (Downing *et al.*, 2001). As the name implies, this region is known for its responses to images of human body parts, and has also been frequently implicated in action observation as well as execution (Astafiev *et al.*, 2004; Kuhn *et al.*, 2011; Liljeholm *et al.*, 2012). Notably, studies assessing the role of the EBA in action execution have employed visually guided actions, such that an arbitrary visual stimulus indicates either the location towards which a movement should be directed (Astafiev *et al.*, 2004), or which one of alternative actions should be performed (Kuhn *et al.*, 2011). It is possible therefore that this area is specifically involved in limb movements that are largely stimulus-driven. In the current study, the S-R condition was designed to encourage a mapping of responses to arbitrary antecedent visual cues by decorrelating actions from sensory-specific outcome features (and indeed from any visual features that were intrinsically related to the goal of maintaining beaker fluids). Thus, one possible explanation for the selective recruitment of EBA by the S-R condition is that this area mediates the detection and mapping of arbitrary stimuli to behavioral responses.

The VMPFC has been implicated in goal-directed performance in numerous human neuroimaging studies, and is a proposed homolog of the rodent prelimbic cortex (Balleine & O’Doherty, 2010), lesions of which disrupt the acquisition, but not the expression, of goal-directed performance (Ostlund & Balleine, 2005). Although we did find selective recruitment of the VMPFC by the R-O condition during learning, this effect did not survive the subtraction of matching control conditions. It is possible, of course, that the process supported by the VMPFC, for example assigning values to subgoals, was elicited by the relevant stimuli in the matching task as well as the instrumental task. Another possibility, given our robust effect of experimental phase in the VMPFC, such that activity was greater during early acquisition than during test in both conditions, is that the goal-directed function supported by the VMPFC was present in both instrumental conditions during early acquisition. Further work is needed to arbitrate between these possibilities.

© 2015 Federation of European Neuroscience Societies and John Wiley & Sons Ltd

Further dissociating the contributions of medial prefrontal areas, we found that activity in a more dorsal medial prefrontal region was greater in the R-O than the S-R condition during test performance, but not during early acquisition. This region, along with the adjacent dorsal anterior cingulate, has been implicated in tasks in which the attainment of goals and rewards requires high levels of cognitive control, such as when monitoring for unfavorable outcomes, and during response conflict and decision uncertainty (Ridderinkhof *et al.*, 2004a,b). The currently observed pattern of activity in this area may reflect a decrease in performance monitoring and outcome evaluation during habitual relative to goal-directed control.

In our novel task the stimulus materials were identical across S-R and R-O conditions, but the instrumental contingencies encouraged participants to attend to different visual features (cues vs. beakers). Although we used a simple match-to-sample task to rule out visual processes involved in selectively attending to such features as a source of any imaging effects, there may be aspects of visual attention that are intrinsically related to instrumental responding. For example, we found greater activity in the R-O than the S-R condition in a posterior ventral region of the IPL during both the learning and test phase; during test, discriminatory activity in this area predicted greater sensitivity to devaluation. The posterior ventral region of the IPL has been proposed to function as a ‘circuit break’ that redirects attention towards behaviorally relevant information that is either exogenously presented (Corbetta & Shulman, 2002; Corbetta *et al.*, 2008; Cabeza *et al.*, 2012) or retrieved into working memory (Cabeza *et al.*, 2012). Importantly, substantial evidence also indicates that the posterior ventral region of the IPL is deactivated when the new information is not relevant to the current task (Shulman *et al.*, 2003). As can be seen in Fig. 3 (top), activity in the posterior ventral region of the IPL appears to be deactivated in both instrumental conditions relevant to the matching control, but markedly more so in the S-R condition. A possible explanation for this pattern of results is that greater suppression was required in the S-R condition because the sensory features of the beaker subgoal, although ultimately irrelevant for response selection, was intimately related to the trial outcome. Indeed, this augmented suppression of sensory-specific outcome features may be a general property of S-R learning, particularly as the retrieval of such features into working memory is considered central to goal-directed encoding (Balleine & Ostlund, 2007).

Consistent with evidence from the rodent literature that goal-directed performance persists in tasks in which alternative actions produce distinct rewards (Colwill & Rescorla, 1985; Holland, 2004), we found that human action selection was more goal-directed in a condition in which instrumental actions obtained unique subgoals. Formally, the degree to which alternative actions yield distinct outcome states can be quantified as the divergence of their outcome probability distributions. In a previous study (Liljeholm *et al.*, 2013), we found that activity in the right anterior IPL (the anterior dorsal supramarginal gyrus) increased with increasing outcome divergence. Likewise, in the current study, training-dependent increases in this area were found in the R-O, but not the S-R, condition, potentially reflecting the incremental increase in divergence as actions became associated with their respective outcome states. The selective recruitment of the IPL by the R-O condition, and the correlation between such discriminatory neural activity and behavioral sensitivity to outcome devaluation, is also consistent with a large body of research implicating this area in various goal-directed processes, including the computation of instrumental contingencies (Seo *et al.*, 2009; Liljeholm *et al.*, 2011), attribution of intent (den Ouden *et al.*, 2005), awareness of agency (Farrer *et al.*, 2008), and mediation of executive function (Friedman & Goldman-Rakic, 1994). Further determination of the exact contributions of the IPL to goal-directed action selection, its role in corollary attentional processes, and in representations of outcome divergence, is an important avenue for future work.

In summary, our results are novel in several critical ways. To our knowledge, we are the first to dissociate the roles of the tail of the caudate nucleus and VS, across learning and test performance, in behavioral insensitivity to outcome devaluation. We are also the first to demonstrate that activity in the IPL, an area that has been previously implicated in several processes closely linked to goal-directed action selection, including the attribution of intent and awareness of agency, predicts sensitivity to outcome devaluation. Finally, we reveal a potential functional homology between the human subgenual and rodent infralimbic cortex in the implementation of habitual control. Taken together, our findings suggest a broad systems division, at the cortical and subcortical levels, between brain areas mediating the acquisition and expression of action–outcome and S-R associations. Notably, a fundamental issue in the search for treatments of behavioral disorders is how to both facilitate the automatization of actions that lead to healthful consequences and abolish well-established deleterious habits. An improved understanding of how distinct neural substrates in the human brain mediate the acquisition vs. expression of habitual control, the aim of the present study, is therefore of significant clinical interest.

Acknowledgements

This work was supported by a grant from the Wellcome Trust and by an NIH grant (DA033077-01) to J.P.O’D. There are no conflicts of interest.

Abbreviations

CST, cluster-size thresholding; EBA, extrastriate body area; IPL, inferior parietal lobule; R-O, Response–Outcome; S-R, Stimulus–Response; SVC, small volume correction; VMPFC, ventromedial prefrontal cortex; VS, ventral striatum.

References

Adams, C.D. (1982) Variations in the sensitivity of instrumental responding to refinforcer devaluation. *Q. J. Exp. Psychol.*, **34**, 77–98.

Adams, C.D. & Dickinson, A. (1981) Instrumental responding following reinforcer devaluation. *Q. J. Exp. Psychol.*, **33**, 109–121.

Ashby, F.G., Turner, B.O. & Horvitz, J.C. (2010) Cortical and basal ganglia contributions to habit learning and automaticity. *Trends Cogn. Sci.*, **14**, 208–215.

Astafiev, S.V., Stanley, C.M., Shulman, G.L. & Corbetta, M. (2004) Extrastriate body area in human occipital cortex responds to the performance of motor actions. *Nat. Neurosci.*, **7**, 542–548.

Balleine, B.W. & Dickinson, A. (1998) Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology*, **37**, 407–419.

Balleine, B.W. & O’Doherty, J.P. (2010) Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology*, **35**, 48–69.

Balleine, B.W. & Ostlund, S.B. (2007) Still at the choice-point: action selection and initiation in instrumental conditioning. *Ann. NY Acad. Sci.*, **1104**, 147–171.

Balsters, J.H. & Ramnani, N. (2011) Cerebellar plasticity and the automation of first-order rules. *J. Neurosci.*, **31**, 2305–2312.

Blaiss, C.A. & Janak, P.H. (2009) The nucleus accumbens core and shell are critical for the expression, but not the consolidation, of Pavlovian conditioned approach. *Behav. Brain Res.*, **200**, 22–32.

Brovelli, A., Nazarian, B., Meunier, M. & Boussaoud, D. (2011) Differential roles of caudate nucleus and putamen during instrumental learning. *NeuroImage*, **57**, 1580–1590.

© 2015 Federation of European Neuroscience Societies and John Wiley & Sons Ltd

Cabeza, R., Ciaramelli, E. & Moscovitch, M. (2012) Cognitive contributions of the ventral parietal cortex: an integrative theoretical account. *Trends Cogn. Sci.*, **16**, 338–352.

Callu, D., Puget, S., Faure, A., Guegan, M. & El Massioui, N. (2007) Habit learning dissociation in rats with lesions to the vermis and the interpositus of the cerebellum. *Neurobiol. Dis.*, **27**, 228–237.

Carelli, R.M. & West, M.O. (1991) Representation of the body by single neurons in the dorsolateral striatum of the awake, unrestrained rat. *J. Comp. Neurol.*, **309**, 231–249.

Carelli, R.M., Wolske, M. & West, M.O. (1997) Loss of lever press-related firing of rat striatal forelimb neurons after repeated sessions in a lever pressing task. *J. Neurosci.*, **17**, 1804–1814.

Colwill, R.M. & Rescorla, R.A. (1985) Instrumental responding remains insensitive to reinforcer devaluation after extensive training. *J. Exp. Psychol. Anim. B.*, **11**, 520–536.

Corbetta, M. & Shulman, G.L. (2002) Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.*, **3**, 201–215.

Corbetta, M., Patel, G. & Shulman, G.L. (2008) The reorienting system of the human brain: from environment to theory of mind. *Neuron*, **58**, 306–324.

Corbit, L.H. & Balleine, B.W. (2011) The general and outcome-specific forms of Pavlovian-instrumental transfer are differentially mediated by the nucleus accumbens core and shell. *J. Neurosci.*, **31**, 11786–11794.

Coutureau, E. & Killcross, S. (2003) Inactivation of the infralimbic prefrontal cortex reinstates goal-directed responding in overtrained rats. *Behav. Brain Res.*, **146**, 167–174.

Daw, N.D., Niv, Y. & Dayan, P. (2005) Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.*, **8**, 1704–1711.

Day, J.J., Roitman, M.F., Wightman, R.M. & Carelli, R.M. (2007) Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens. *Nat. Neurosci.*, **10**, 1020–1028.

Dickinson, A., Nicholas, D.J. & Adams, C.D. (1983) The effect of the instrumental training contingency on susceptibility to reinforcer devaluation. *Q. J. Exp. Psychol.-B.*, **35**, 35–51.

Downing, P.E., Jiang, Y., Shuman, M. & Kanwisher, N. (2001) A cortical area selective for visual processing of the human body. *Science*, **293**, 2470–2473.

Doya, K., Samejima, K., Katagiri, K. & Kawato, M. (2002) Multiple model-based reinforcement learning. *Neural Comput.*, **14**, 1347–1369.

Doyon, J., Laforce, R. Jr., Bouchard, G., Gaudreau, D., Roy, J., Poirier, M., Bedard, P.J., Bedard, F. & Bouchard, J.P. (1998) Role of the striatum, cerebellum and frontal lobes in the automatization of a repeated visuomotor sequence of movements. *Neuropsychologia*, **36**, 625–641.

Doyon, J., Song, A.W., Karni, A., Lalonde, F., Adams, M.M. & Ungerleider, L.G. (2002) Experience-dependent changes in cerebellar contributions to motor sequence learning. *Proc. Natl. Acad. Sci. USA*, **99**, 1017–1022.

Draganski, B., Kherif, F., Kloppel, S., Cook, P.A., Alexander, D.C., Parker, G.J., Deichmann, R., Ashburner, J. & Frackowiak, R.S. (2008) Evidence for segregated and integrative connectivity patterns in the human Basal Ganglia. *J. Neurosci.*, **28**, 7143–7152.

Drevets, W.C., Savitz, J. & Trimble, M. (2008) The subgenual anterior cingulate cortex in mood disorders. *CNS Spectr.*, **13**, 663–681.

Esterman, M., Tamber-Rosenau, B.J., Chiu, Y.C. & Yantis, S. (2010) Avoiding non-independence in fMRI data analysis: leave one subject out. *NeuroImage*, **50**, 572–576.

Farrer, C., Frey, S.H., Van Horn, J.D., Tunik, E., Turk, D., Inati, S. & Grafton, S.T. (2008) The angular gyrus computes action awareness representations. *Cereb. Cortex*, **18**, 254–261.

Floyer-Lea, A. & Matthews, P.M. (2004) Changing brain networks for visuomotor control with increased movement automaticity. *J. Neurophysiol.*, **92**, 2405–2412.

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.

Friedman, H.R. & Goldman-Rakic, P.S. (1994) Coactivation of prefrontal cortex and inferior parietal cortex in working memory tasks revealed by 2DG functional mapping in the rhesus monkey. *J. Neurosci.*, **14**, 2775–2788.

Gabbott, P.L., Dickie, B.G., Vaid, R.R., Headlam, A.J. & Bacon, S.J. (1997) Local-circuit neurones in the medial prefrontal cortex (areas 25, 32 and 24b) in the rat: morphology and quantitative distribution. *J. Comp. Neurol.*, **377**, 465–499.

Glascher, J. (2009) Visualization of group inference data in functional neuroimaging. *Neuroinformatics*, **7**, 73–82.

Greicius, M.D., Flores, B.H., Menon, V., Glover, G.H., Solvason, H.B., Kenna, H., Reiss, A.L. & Schatzberg, A.F. (2007) Resting-state functional connectivity in major depression: abnormally increased contributions from subgenual cingulate cortex and thalamus. *Biol. Psychiat.*, **62**, 429–437.

Haddon, J.E. & Killcross, S. (2011) Inactivation of the infralimbic prefrontal cortex in rats reduces the influence of inappropriate habitual responding in a response-conflict task. *Neuroscience*, **199**, 205–212.

Hampton, A.N., Bossaerts, P. & O'Doherty, J.P. (2006) The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *J. Neurosci.*, **26**, 8360–8367.

Hikosaka, O., Kim, H.F., Yasuda, M. & Yamamoto, S. (2014) Basal ganglia circuits for reward value-guided behavior. *Annu. Rev. Neurosci.*, **37**, 289–306.

Holland, P.C. (2004) Relations between Pavlovian-instrumental transfer and reinforcer devaluation. *J. Exp. Psychol. Anim. B.*, **30**, 104–117.

Jog, M.S., Kubota, Y., Connolly, C.I., Hillegaart, V. & Graybiel, A.M. (1999) Building neural representations of habits. *Science*, **286**, 1745–1749.

Johansen-Berg, H., Gutman, D.A., Behrens, T.E., Matthews, P.M., Rushworth, M.F., Katz, E., Lozano, A.M. & Mayberg, H.S. (2008) Anatomical connectivity of the subgenual cingulate region targeted with deep brain stimulation for treatment-resistant depression. *Cereb. Cortex*, **18**, 1374–1383.

Keedwell, P.A., Drapier, D., Surguladze, S., Giampietro, V., Brammer, M. & Phillips, M. (2010) Subgenual cingulate and visual cortex responses to sad faces predict clinical outcome during antidepressant treatment for depression. *J. Affect. Disorders*, **120**, 120–125.

Killcross, S. & Coutureau, E. (2003) Coordination of actions and habits in the medial prefrontal cortex of rats. *Cereb. Cortex*, **13**, 400–408.

Kim, H.F. & Hikosaka, O. (2013) Distinct basal ganglia circuits controlling behaviors guided by flexible and stable values. *Neuron*, **79**, 1001–1010.

Kim, H.F., Ghazizadeh, A. & Hikosaka, O. (2014) Separate groups of dopamine neurons innervate caudate head and tail encoding flexible and stable value memories. *Front. Neuroanatomy*, **8**, 120.

Kuhn, S., Keizer, A., Rombouts, S.A. & Hommel, B. (2011) The functional and neural mechanism of action preparation: roles of EBA and FFA in voluntary action control. *J. Cognitive Neurosci.*, **23**, 214–220.

Lang, C.E. & Bastian, A.J. (2002) Cerebellar damage impairs automaticity of a recently practiced movement. *J. Neurophysiol.*, **87**, 1336–1347.

Lee, S.W., Shimojo, S. & O'Doherty, J.P. (2014) Neural computations underlying arbitration between model-based and model-free learning. *Neuron*, **81**, 687–699.

Liljeholm, M., Tricomi, E., O'Doherty, J.P. & Balleine, B.W. (2011) Neural correlates of instrumental contingency learning: differential effects of action-reward conjunction and disjunction. *J. Neurosci.*, **31**, 2474–2480.

Liljeholm, M., Molloy, C.J. & O'Doherty, J.P. (2012) Dissociable brain systems mediate vicarious learning of stimulus-response and action-outcome contingencies. *J. Neurosci.*, **32**, 9878–9886.

Liljeholm, M., Wang, S., Zhang, J. & O'Doherty, J.P. (2013) Neural correlates of the divergence of instrumental probability distributions. *J. Neurosci.*, **33**, 12519–12527.

Maldjian, J.A., Laurienti, P.J., Kraft, R.A. & Burdette, J.H. (2003) An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *NeuroImage*, **19**, 1233–1239.

Matthews, S., Simmons, A., Strigo, I., Gianaros, P., Yang, T. & Paulus, M. (2009) Inhibition-related activity in subgenual cingulate is associated with symptom severity in major depression. *Psychiat. Res.*, **172**, 1–6.

Milad, M.R. & Quirk, G.J. (2002) Neurons in medial prefrontal cortex signal memory for fear extinction. *Nature*, **420**, 70–74.

O'Doherty, J.P., Deichmann, R., Critchley, H.D. & Dolan, R.J. (2002) Neural responses during anticipation of a primary taste reward. *Neuron*, **33**, 815–826.

Ongur, D. & Price, J.L. (2000) The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cereb. Cortex*, **10**, 206–219.

Ongur, D., Ferry, A.T. & Price, J.L. (2003) Architectonic subdivision of the human orbital and medial prefrontal cortex. *J. Comp. Neurol.*, **460**, 425–449.

Orban, P., Peigneux, P., Lungu, O., Albouy, G., Breton, E., Laberenne, F., Benali, H., Maquet, P. & Doyon, J. (2010) The multifaceted nature of the relationship between performance and brain activity in motor sequence learning. *NeuroImage*, **49**, 694–702.

Ostlund, S.B. & Balleine, B.W. (2005) Lesions of medial prefrontal cortex disrupt the acquisition but not the expression of goal-directed learning. *J. Neurosci.*, **25**, 7763–7770.

den Ouden, H.M., Frith, U. & Frith, C.S.J.B. (2005) Thinking about intentions. *NeuroImage*, **28**, 787–796.

Poldrack, R.A. & Gabrieli, J.D. (2001) Characterizing the neural mechanisms of skill learning and repetition priming: evidence from mirror reading. *Brain*, **124**, 67–82.

Poldrack, R.A., Sabb, F.W., Foerde, K., Tom, S.M., Asarnow, R.F., Bookheimer, S.Y. & Knowlton, B.J. (2005) The neural correlates of motor skill automaticity. *J. Neurosci.*, **25**, 5356–5364.

Rhodes, S.E. & Killcross, S. (2004) Lesions of rat infralimbic cortex enhance recovery and reinstatement of an appetitive Pavlovian response. *Learn. Memory*, **11**, 611–616.

Rhodes, S.E. & Killcross, A.S. (2007) Lesions of rat infralimbic cortex enhance renewal of extinguished appetitive Pavlovian responding. *Eur. J. Neurosci.*, **25**, 2498–2503.

Ridderinkhof, K.R., van den Wildenberg, W.P., Segalowitz, S.J. & Carter, C.S. (2004a) Neurocognitive mechanisms of cognitive control: the role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning. *Brain Cognition*, **56**, 129–140.

Ridderinkhof, K.R., Ullsperger, M., Crone, E.A. & Nieuwenhuis, S. (2004b) The role of the medial frontal cortex in cognitive control. *Science*, **306**, 443–447.

Santini, E., Quirk, G.J. & Porter, J.T. (2008) Fear conditioning and extinction differentially modify the intrinsic excitability of infralimbic neurons. *J. Neurosci.*, **28**, 4028–4036.

Seo, H., Barraclough, D.J. & Lee, D. (2009) Lateral intraparietal cortex and reinforcement learning during a mixed-strategy game. *J. Neurosci.*, **29**, 7278–7289.

Shulman, G.L., McAvoy, M.P., Cowan, M.C., Astafiev, S.V., Tansy, A.P., d'Avossa, G. & Corbetta, M. (2003) Quantitative analysis of attention and detection signals during visual search. *J. Neurophysiol.*, **90**, 3384–3397.

Smith, K.S. & Graybiel, A.M. (2013) A dual operator view of habitual behavior reflecting cortical and striatal dynamics. *Neuron*, **79**, 361–374.

Smith, K.S., Virkud, A., Deisseroth, K. & Graybiel, A.M. (2012) Reversible online control of habitual behavior by optogenetic perturbation of medial prefrontal cortex. *Proc. Natl. Acad. Sci. USA*, **109**, 18932–18937.

Tanaka, S.C., Balleine, B.W. & O'Doherty, J.P. (2008) Calculating consequences: brain systems that encode the causal effects of actions. *J. Neurosci.*, **28**, 6750–6755.

Tang, C., Pawlak, A.P., Prokopenko, V. & West, M.O. (2007) Changes in activity of the striatum during formation of a motor habit. *Eur. J. Neurosci.*, **25**, 1212–1227.

Tricomi, E., Balleine, B.W. & O'Doherty, J.P. (2009) A specific role for posterior dorsolateral striatum in human habit learning. *Eur. J. Neurosci.*, **29**, 2225–2232.

Turner, R.S., McCairn, K., Simmons, D. & Bar-Gad, I. (2005) Sequential motor behavior and the basal ganglia. In Bolam, J.P., Ingham, C.A. & Magill, P.J. (Eds), *The Basal Ganglia VIII (Advances in Behavioral Biology)*. Springer, New York, pp. 563–574.

Ungerleider, L.G., Doyon, J. & Karni, A. (2002) Imaging brain plasticity during motor skill learning. *Neurobiol. Learn. Mem.*, **78**, 553–564.

Valentin, V.V., Dickinson, A. & O'Doherty, J.P. (2007) Determining the neural substrates of goal-directed learning in the human brain. *J. Neurosci.*, **27**, 4019–4026.

Ward, B.D. (2000) Simultaneous inference for fMRI data [Computer software manual]. AFNI 3dDeconvolve Documentation, Medical College of Wisconsin. [Internet] Available http://afni.nimh.nih.gov/afni/doc/manual/AlphaSim.

de Wit, S., Watson, P., Harsay, H.A., Cohen, M.X., van de Vijver, I. & Ridderinkhof, K.R. (2012) Corticostriatal connectivity underlies individual differences in the balance between habitual and goal-directed action control. *J. Neurosci.*, **32**, 12066–12075.

Wunderlich, K., Dayan, P. & Dolan, R.J. (2012) Mapping value based planning and extensively trained choice in the human brain. *Nat. Neurosci.*, **15**, 786–791.

Yamamoto, S., Kim, H.F. & Hikosaka, O. (2013) Reward value-contingent changes of visual responses in the primate caudate tail associated with a visuomotor skill. *J. Neurosci.*, **33**, 11227–11238.

Yin, H.H., Knowlton, B.J. & Balleine, B.W. (2004) Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *Eur. J. Neurosci.*, **19**, 181–189.

Yin, H.H., Knowlton, B.J. & Balleine, B.W. (2005) Blockade of NMDA receptors in the dorsomedial striatum prevents action-outcome learning in instrumental conditioning. *Eur. J. Neurosci.*, **22**, 505–512.

© 2015 Federation of European Neuroscience Societies and John Wiley & Sons Ltd