NIH Public Access
Author Manuscript
Eur J Neurosci. Author manuscript; available in PMC 2010 June 1.

Published in final edited form as:
Eur J Neurosci. 2009 June ; 29(11): 2225–2232. doi:10.1111/j.1460-9568.2009.06796.x.

A specific role for posterior dorsolateral striatum in human habit learning

Elizabeth Tricomi[1], **Bernard W. Balleine**[2], and **John P. O'Doherty**[3,4,5]

[1] Department of Psychology, 353 Smith Hall, Rutgers University, Newark, NJ 07102 [2] Department of Psychology and the Brain Research Institute, 1285 Franz Hall, University of California, Los Angeles, CA 90024 [3] Division of the Humanities and Social Sciences, California Institute of Technology, 1200 East California Blvd., Pasadena, CA 91125 [4] Computation and Neural Systems Program, California Institute of Technology, 1200 East California Blvd., Pasadena, CA, 91125 [5] School of Psychology and Trinity College Institute of Neuroscience, Trinity College, Dublin, Ireland

Abstract

Habits are characterized by an insensitivity to their consequences and, as such, can be distinguished from goal-directed actions. The neural basis of the development of demonstrably outcome insensitive habitual actions in humans has not been previously characterized. In this experiment, we show that extensive training on a free-operant task reduces the sensitivity of participants' behavior to a reduction in outcome value. Analysis of functional magnetic resonance imagine (fMRI) data acquired during training revealed a significant increase in task-related cue sensitivity in a right posterior putamen/globus pallidus region as training progressed. These results provide evidence for a shift from goal-directed to habit-based control of instrumental actions in humans, and suggest that cue-driven activation in a specific region of dorsolateral posterior putamen may contribute to the habitual control of behavior in humans.

Keywords

reward; instrumental conditioning; fMRI; putamen; basal ganglia

Awareness of the consequences of one's actions allows behavior to be flexibly modified as the value of those consequences changes. In contrast, the development of habits allows responses to be efficiently executed, freeing up valuable cognitive resources. Research in animals has shown that whereas newly acquired actions are goal-directed and sensitive to changes in outcome value, with continued invariant training their performance can become more reflexive or habitual and controlled by antecedent stimuli rather than their consequences. If a stimulus-response (S-R) habit has been formed, a cue will elicit its associated response regardless of outcome value (Dickinson & Balleine, 2002).

Research in rodents has demonstrated that different neural systems contribute to these two types of learning, with the prefrontal cortex and the dorsomedial striatum subserving goal-directed action, and the dorsolateral striatum (DLS) supporting habit-based behavior (Balleine & Dickinson, 1998a; Killcross & Coutureau, 2003; Yin *et al*., 2004; Yin *et al*., 2005). Although some neuroimaging studies have supported a role for prefrontal cortex and

Corresponding author address: Elizabeth Tricomi Dept. of Psychology, Rutgers University 353 Smith Hall, 101 Warren Street Newark, NJ 07102 Phone: (973) 353-5440 ×3956 Fax: (973) 353-1171 etricomi@psychology.rutgers.edu.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

anterior caudate nucleus in goal-directed behavior in humans (O'Doherty *et al*., 2004; Tricomi *et al*., 2004; Valentin *et al*., 2007; Tanaka *et al*., 2008), much less is known about the specific neural circuitry responsible for outcome insensitive habitual behavior in the human brain. A number of human studies have examined procedural learning, which is often described as "habit"-like (Knowlton *et al*., 1996; Bayley *et al*., 2005; Foerde *et al*., 2006), but the habitual nature of this form of responding has been inferred and has never been shown in humans to be outcome insensitive; i.e., it has never been established whether actions determined by this form of learning are sensitive to outcome devaluation. This study had two goals: firstly, we aimed to determine whether instrumental responding in human subjects can be rendered habitual; i.e., whether it can be rendered outcome insensitive so that it will persist even after an outcome is no longer valued, and; secondly, we aimed to uncover the neural mechanisms that contribute to the development and/or control of habitual behavior in humans, with a particular emphasis on the DLS, known to contribute to this process in rodents (Yin *et al*., 2004).

We developed a free-operant task inspired by experimental paradigms used in the animal literature (e.g., Dickinson *et al*., 1995), in which human subjects were either given relatively little training or were overtrained to press a button, with a rewarding outcome delivered on a variable-interval (VI) reinforcement schedule. Performance was then assessed in an extinction test after outcome devaluation. Responding was self-paced, an approach commonly used in animal research but that contrasts with the trial-based paradigms typically used in human and monkey studies in which single responses are cued on a trial-by-trial basis. This design permitted the use of response rate as a direct measure of performance and as an index of sensitivity to changes in the incentive value of the outcome as well as allowing greater comparability between the findings reported here and those reported in studies on habitual learning in the animal literature.

Materials and Methods

Participants

Thirty-two healthy participants participated in the experiment (21 females, 11 males; mean age: 22.0 ± 2.7 SD, range: 19-28). All participants were prescreened to ensure that they were not dieting and that they enjoyed eating chocolate and corn chips. Since the experiment involved food consumption, the eating attitudes test (EAT-26) (Garner *et al*., 1982) was administered prior to the experiment, which indicated no eating disorders in any of the participants (mean score: 3.0 ± 3.6 SD; range: 0-17). All scores were under the 20-point cutoff. However, since some studies have suggested a lower threshold for screening (Orbitello *et al*., 2006), we examined whether the experimental data from the two subjects with a score of over 11 (one from each group) were outliers. They were not, so data from these subjects were included in the analyses. Participants were asked to fast for at least 6 h prior to each scan, although they were allowed to drink water. All participants gave written informed consent and the study conforms with The Code of Ethics of the World Medical Association (Declaration of Helsinki, 1964). The Institutional Review Board at the California Institute of Technology approved the study.

Experimental procedure

Participants were scanned during training on a free-operant task (Fig. 1) in which responses to fractal cues were rewarded on a variable interval (VI) schedule with M&M's® (Mars, McLean, VA, USA) and Fritos® (Frito-Lay, Inc., Plano, TX, USA), to be consumed following the scan. One group of participants (n=16) performed two 8-minute training sessions on one day (1-day group), whereas a second group of participants (n=16) performed

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

four training sessions each day for three days (3-day group). Therefore the 3-day group had six times as much training on the task as the 1-day group.

Stimulus presentation and behavioral data acquisition were implemented in Matlab (The Mathworks Inc., Natick, MA, USA) with the Cogent 2000 toolbox (Wellcome Department of Imaging Neuroscience, London, UK). Each session was divided into 12 task blocks (20-40 s each) and 8 rest blocks (20 s each). Unlike most fMRI studies, we did not used a trial-based paradigm, but instead employed a free operant paradigm, in which responses are self-paced. During the task blocks, a fractal image was shown on the screen, along with a schematic indicating which button to press. Thus the onset and offset of the fractal image display indicated the start and end of the block. Participants were instructed to press the indicated button as often as they liked; after each button press either a gray circle briefly appeared (50 ms), indicating no reward, or a picture of an M&M or Frito appeared (1000 ms), indicating a food reward corresponding to the picture. Only presses of the indicated button led to the display of the gray circle or food picture; if a different button was pressed, the display did not change. Rewards were delivered on a VI 10-s schedule (i.e., each second there was a 0.1 chance that a reward would become available, and the reward was delivered upon the first button press following reward availability. Therefore, a reward became available on average every 10 s, with an equal probability of becoming available each second). A VI schedule was chosen because animals have been found to acquire demonstrably habitual behavior more rapidly when trained with a VI schedule than with a variable-ratio schedule (Dickinson *et al*., 1983). The accumulated earnings were consumed by the participants following the scan. Different fractals and response keys were paired with the two outcomes, and these stimulus-response-outcome pairings remained consistent throughout the experiment (although they were counterbalanced across participants). A third fractal indicated a rest block, during which participants were instructed not to respond. The block order was pseudorandomized, with no block type occurring twice in a row.

Following the final session of training, one of the two food outcomes was devalued through selective satiation (Rolls *et al*., 1983; Colwill & Rescorla, 1985; Balleine & Dickinson, 1998b; O'Doherty *et al*., 2000; Gottfried *et al*., 2003), in which participants were asked to eat that food until it was no longer pleasant to them. The food chosen for devaluation was counterbalanced across participants. To test the effects of the devaluation procedure on behavior, participants were placed back into the scanner for a 3-min extinction test. This provides an explicit test for the presence of habitual behavior. If behavior remains goal-directed, participants should respond less for the food they no longer find pleasant than for the still valued food; however, if behavior has become habitual, the fractal cues should elicit responding irrespective of the outcome value. The subjects were given the same instructions and the extinction test was implemented in the same manner as for the training sessions; however, no rewards were actually delivered during extinction. Testing during extinction ensures that behavior is based on previously acquired associations, independent of any potential impact of delivery of the devalued outcome itself. The extinction test was kept short because subjects eventually realize no rewards are being delivered. The extinction phase was conducted inside the scanner to keep the extinction context similar to the testing context, but because the extinction phase is only 3 min, there is insufficient usable data from the extinction phase to be used in the fMRI analysis. The sole purpose of the extinction phase was to provide a behavioral assay of the degree of habitization after training in the two groups. Additionally, Likert-scale ratings of hunger (1, very full; 10, very hungry) and pleasantness of each food (−5, very unpleasant; 5, very pleasant) were obtained prior to each day's training session and following the devaluation procedure.

One subject in the 3-day group was treated as an outlier and excluded from analysis because the response rate difference between conditions during extinction was greater than 2

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

standard deviations from the group mean. It should, however, be noted that all of the results discussed here remained statistically significant with all of the subjects included, and figures depicting the results with all the subjects can be found in the supporting online information (Supporting Figures S1-S4). The final group numbers for the statistical analysis were Group 1-day: n=16; Group 3-day n=15.

fMRI data acquisition

A 3 Tesla Siemens (Erlangen, Germany) Trio scanner and an 8-channel phased array coil was used to acquire both high-resolution T1-weighted structural images (1 × 1 × 1 mm) for anatomical localization and T2*-weighted echo planar images (45 slices, 3 × 3 × 3 mm voxels, TR = 2.65 s, TE = 30 ms, flip angle = 80°, FoV = 192 × 192 mm, slice gap = 0 mm) with BOLD (Blood Oxygen Level Dependent) contrast. Each image was acquired in an oblique orientation of 30° to the anterior commissure-posterior commissure (AC-PC) axis, which reduces signal dropout in the ventral prefrontal cortex relative to AC-PC aligned images (Deichmann *et al*., 2003).

Data Analysis

To determine whether the devaluation procedure differentially affected response rates for the two groups of participants, a repeated measures ANOVA was performed on measures of the change in response rate between the extinction test and the last session of training, with subject as a random factor, training (1-day vs. 3-day) as a between-subjects factor, and devaluation (valued vs. devalued outcome) as a within-subjects factor. Post-hoc t-tests were performed to determine whether there were significant differences between specific conditions. To confirm that the devaluation procedure was effective in selectively reducing pleasantness for the devalued outcome, t-tests were performed on the difference in pleasantness ratings from the beginning of the day's training to after devaluation for the valued versus the devalued outcome. T-tests were also used to compare the changes in hunger and food pleasantness ratings in the two groups of participants.

We used SPM5 (Wellcome Department of Imaging Neuroscience, Institute of Neurology, London, UK) for preprocessing and analysis of the fMRI data. The images were slice-time corrected to the middle of the image acquisitions, realigned to the first volume to correct for subject motion, spatially transformed to match the Montreal Neurological Institute EPI template, and spatially smoothed with a Gaussian kernel (8 mm, full width at half maximum). We also applied intensity normalization and high-pass filtering (filter width = 200 s) to the imaging data.

The general linear model was used to generate voxelwise statistical parametric maps from the functional data. For each participant, we constructed an fMRI design matrix by modeling the following regressors for each training session: "task onset" (modeled as a stick function at the onset of each task block), "rest onset" (a stick function at the onset of each rest block), "task" (modeled as a series of 1-s events within each task block), "ramp" (a parametric modulator of the task regressor indicating the number of seconds since the previous reward presentation or start of the block) and "rest" (which modeled the rest blocks), and "reward" (a stick function at the onset of each reward presentation). The fMRI data from the extinction test were not included in the analysis.

The regressors were serially orthogonalized, in the order listed above, and convolved with a canonical hemodynamic response function. Regressors of no interest were also generated using the realignment parameters from the image preprocessing to further correct for residual subject motion. The design matrix was then regressed against the fMRI data in order to generate parameter estimates for each subject. The parameter estimates generated

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

from this 1st-level analysis were entered into a random-effects group analysis, and linear contrasts were used to identify significant effects at the group level. For anatomical localization, the statistical maps were rendered on the average of all subjects' high-resolution T1-weighted structural images.

Because the devaluation and extinction test procedures could only be performed once per subject without potentially biasing future behavior, it was necessary to use a between-subjects design to examine the effects of training on the motivational control of behavior. However, in addition to a between-subjects analysis, we were able to employ a within-subjects design to examine training-related changes in brain activation, hence we acquired fMRI data throughout all the training sessions. Inter-subject variability in BOLD response is often quite high, so a within-subjects design (for the fMRI component of the study) overcomes this problem, allowing analyses to be performed that have adequate statistical power, given the limited sample sizes that are practical for fMRI experiments.

To identify brain regions whose sensitivity to task-related cues changed over training, we examined a contrast of [task onset — rest onset] for the last 2 sessions of training in the 3-day group compared to the 2 sessions of training in the 1-day group. We also compared the last 2 versus first 2 sessions of training within the 3-day group, as well as within-day learning effects through a contrast of task onset versus rest onset increasing across the four sessions of each day. Since we had an a priori hypothesis that the DLS, thought to correspond to the putamen in primates, would play a role in S-R habit learning, we used the WFU PickAtlas toolbox to apply an anatomical mask of the putamen (dilation = 1) to the data (Maldjian *et al*., 2003; Maldjian *et al*., 2004). A small volume correction for the masked area was then performed, with a corrected significance threshold of $P < 0.05$.

In addition, for each group, brain regions displaying a phasic response to reward presentation were identified by their sensitivity to the reward contrast. Finally, we sought to identify brain regions displaying goal sensitivity early in training; such regions should show increased activation as anticipation of an upcoming reward increases. The data from the first two sessions from the participants in both groups were included in a random-effects analysis, and brain regions in which the BOLD signal ramped up until the next reward presentation were identified by their sensitivity to the ramp modulator. For all contrasts, a voxel-wise significance threshold of $p < 0.001$ was used, along with a contiguity threshold of five voxels as a precaution against type-1 errors (Forman *et al*., 1995). To visualize the effects of training, plots of contrast values for each training session were made in regions of interest that met our significance threshold. If the region encompassed multiple anatomical regions with separate significance peaks, the threshold was raised until the contrast values for the region of interest could be plotted separately.

Results

The 1-day group of subjects was scanned during performance of two 8-min training sessions on one day, whereas the 3-day group was scanned during the performance of four training sessions a day for three days. During the last session of training, there were no significant differences in response rates between groups ($t_{(29)} = 0.8$, $P > 0.05$) or when responding for the two food rewards ($t_{(30)} = 1.6$, $P > 0.05$). Analysis of Likert-scale ratings confirm that, following the devaluation procedure, pleasantness ratings of the devalued food decreased significantly more than for the valued food ($t_{(15)} = -5.5$ for 1-day group, $t_{(14)} = -5.1$ for 3-day group; $P < 0.001$). This cannot be attributed to a general decrease in pleasantness over extended training for the 3-day subjects, because the pleasantness ratings for the valued food did not differ significantly between groups ($t_{(29)} = 0.6$, $p > 0.05$). There were no significant differences between groups in the change in food pleasantness ratings ($t_{(29)} = -1.0$ for

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

devalued outcome, $t_{(29)} = 0.01$ for valued outcome; $P > 0.05$) or hunger ratings ($t_{(29)} = 0.7$, p > 0.05). Therefore, although the ratings were subjective and thus could potentially have been influenced by factors such as demand characteristics of the instructions, the ratings made by the two groups were not differentially affected by the devaluation procedure and are consistent with the interpretation that the subjects became sated on the devalued but not the valued food.

During the extinction test, participants in the 1-day group reduced their response rates during presentation of the cue linked to the devalued food, indicating that their behavior was goal-directed, whereas participants in the 3-day group continued to respond for the devalued food, indicating that their behavior had become insensitive to changes in outcome value over the course of training. This produced a significant training by devaluation interaction ($F_{(1,29)} = 7.9$, $P < 0.05$; Fig. 2). Post-hoc t-tests confirm that response rates for the devalued outcome decreased significantly more than for the valued outcome for the 1-day group ($t_{(15)} = 3.6$, $P < 0.01$), whereas there was no significant difference for the 3-day group ($t_{(14)} = 0.2$, $P > 0.05$). Furthermore, during the extinction test, response rates for the devalued outcome decreased significantly more for the 1-day group than the 3-day group ($t_{(29)} = 2.3$, $P < 0.01$), but there was not a significant group difference for the valued outcome ($t_{(29)} = 0.1$, $P > 0.05$). Thus, whereas the behavior of the two groups was similar prior to devaluation, the devaluation procedure had a differential effect on the response rates of the two groups for the devalued outcome.

We used the general linear model to analyze the fMRI data. In addition to modeling the task and rest blocks, we modeled the onset of each task block and the onset of each rest block as phasic events, to capture any changes in BOLD signal specifically related to the onset of the fractal cues. According to associative learning theory, habit learning should be marked by increasing sensitivity to antecedent stimuli linked with a particular response. A between-subjects comparison of [task onset — rest onset] for the last 2 sessions of training of the 3-day group versus the 2 sessions of training of the 1-day group showed one cluster in the left superior temporal lobe that was significantly greater for the 3-day group at a threshold of $P < 0.001$ ($t_{(29)} > 3.4$). Our within-subjects analysis of the last two sessions of training versus the first two sessions in the 3-day group, however, revealed several significant voxel clusters (Supporting Table S1), including a region within the DLS, in the right posterior putamen/globus pallidus (Fig. 3). This contrast was significantly greater during the final two sessions of training on the last day compared to the first two sessions of the first day ($t_{(14)} > 3.8$, $P < 0.001$), and survived a small volume correction for the area within an anatomical mask of the putamen ($P(\text{cor}) = 0.036$). Based on the behavioral results, behavior should still be goal-directed during the first two sessions of training. For illustration purposes, we have also graphed the contrast values for the 1-day group in this region; these values are similar to those from the first two sessions of the 3-day group. Our analysis of within-day training effects revealed a similar DLS region which showed an increasing response to the onset of task blocks relative to the onset of rest blocks over the course of each day of training ($t_{(14)} > 3.8$, $P < 0.001$, $P(\text{cor}) = 0.003$). These results suggest that there is an increase in task-related stimulus-driven activation with training, which may reflect an increasing contribution of the DLS to governing behavior as the S-R habit develops.

To identify brain regions sensitive to reward presentation, we modeled the presentation of rewards as phasic events. The left nucleus accumbens displayed a significant response to reward presentation for both groups ($t_{(15)} > 3.7$ for 1-day group, $t_{(14)} > 3.8$ for 3-day group, $P < 0.001$, Fig. 4, Tables S2 and S3). The magnitude of this response stayed relatively consistent over the course of training.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Finally, we reasoned that brain regions involved in goal-directed action should display increased activation as anticipation of an upcoming reward increases. We therefore analyzed the first two sessions of fMRI data from both groups to identify brain regions displaying a sensitivity to a "ramp" modulator, which indicated the number of seconds since the previous reward presentation or start of the block (i.e., regions which "ramped up" until reward presentation). The ventromedial prefrontal cortex (vmPFC) displayed a significant effect ($t_{(30)} > 3.4$, $P < 0.001$; Fig. 5, Table S4). We plotted the contrast value in this region over each session in the 3-day group; however, we did not observe any clear pattern of change in this effect over time.

Discussion

Our findings complement the animal literature on the neural basis of habit learning and provide evidence for a habit learning system in humans. Although we all have anecdotal experience of performing outcome-inappropriate cue-driven behavior (e.g., stepping out of an elevator when the doors open, although it has stopped on the wrong floor), one might expect that humans would be able to suppress habitual tendencies more easily than other animals and would not repeatedly make outcome-inappropriate responses in a free-operant task. Our behavioral data show, however, a clear experience-dependent shift from goal-directed to habitual behavior in a free-operant task in humans. After minimal training, participants reduced their response rates during presentation of the fractal linked with the food they no longer wanted, whereas their response rates remained high during presentation of the fractal linked with the food they still found pleasant. After more extensive training, this outcome sensitivity was not present; response rates did not differ significantly during presentation of fractals whether linked with the valued or the devalued outcome.

In our experiment, we found a region in the posterior putamen extending into the globus pallidus that showed increasingly greater activation with experience to the onset of task-related stimuli relative to the onset of rest-related stimuli. In other words, this region became increasingly sensitive to stimuli that were associated with a particular behavioral response, consistent with a potential role in S-R learning. Based on our behavioral findings, 12 sessions of training on our task was enough to elicit habitual behavior, whereas after only two sessions of training, actions were still goal-directed. Our imaging results in this region correspond well with these behavioral results, in that this region showed a significant difference in task vs. rest cue sensitivity between the last two and the first two sessions of training, suggesting that the posterior putamen/globus pallidus region may play a central role in the development and/or control of habitual behavior in humans.

We also identified a similar area showing a within-day increase in task versus rest cue sensitivity. As Figure 3B shows, during the last session of the second day of training, this sensitivity is almost as strong as during the last session of the last day of training. This sensitivity appears to be diminished at the beginning of the third day of training, and then it increases again over the course of training that day. This may reflect a potential resurgence in goal-directed responding relative to habitual responding at the beginning of task performance each day. However, the fact that there is also a significant difference between the first two sessions of training on the first day and the last two sessions on the last day indicates there is also a cumulative effect of multiple days of training on activation in this region, consistent with the notion that this area becomes more involved after successive days of training, mirroring the behavioral development of habits.

These data also underscore the point that the transition from goal-directed to habitual control of behavior is highly dynamic and that the early phase of the habit learning process occurs even while behavior is still demonstrably goal-directed (Graybiel, 2008). As these results

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

make clear, it is not the case that the DLS is suddenly engaged at the moment that behavior becomes habitual. Rather, the recruitment of the DLS, and the degree to which S-R associations influence performance, increases gradually with training (Balleine & Ostlund, 2007). Note, for example, that there is a slight (nonsignificant) increase in sensitivity of the DLS to the block onsets in the second session compared to the first session of training for the 1-day group. This small increase, however, is not enough to result in habitual behavior following devaluation; only with extended training and further increases in the sensitivity of the DLS to task-relevant cues is there a significant shift toward responding habitually following outcome devaluation.

The results of this study suggest that S-R habit learning may be mediated separately from goal-directed learning; in contrast to habits, goal-directed learning is commonly thought to depend on a process of response-outcome association (Colwill & Rescorla, 1985; Dickinson & Balleine, 2002). This is in line with evidence from research on rodents that the development of S-R habits, but not response-outcome associations, relies on the DLS, which corresponds to the dorsal putamen in humans (Yin *et al*., 2004). The region we identified lies on the border of the putamen and globus pallidus, two basal ganglia subregions which are highly interconnected (Spooren *et al*., 1996), and which are thought to play an important role in the “motor loop” of the cortico-basal ganglia-thalamo-cortical pathway (Parent & Hazrati, 1995). Although this connectivity puts this area in a prime position to influence behavioral responses, the results of this study suggest that its role extends beyond motor control to a role in building up S-R associations. Indeed, the portion of the putamen that we found to show increasing sensitivity to response-linked cues is different from a left-lateralized, more anterior portion of the putamen that can be identified by a simple task versus rest comparison.

In contrast to the DLS, the vmPFC has been implicated in governing goal-directed action in humans (Valentin *et al*., 2007). This region may play a role in supporting goal-directed behavior by representing the value of the upcoming outcome (Schoenbaum *et al*., 1998; Tanaka *et al*., 2004; Daw *et al*., 2006; Hampton *et al*., 2006; Kim *et al*., 2006; Roesch & Olson, 2007). In our experiment, we found that the vmPFC shows activation that ramps up from the block onset or previous reward until the next reward is presented, which is consistent with the idea that this region is involved in anticipation of an upcoming reward. Although performance of well-trained actions on VI schedules is not controlled by reward expectation within the interval, the role of reward expectation (and of vmPFC) in performance in undertrained actions suggests that this ramping may play a role in goal-directed but not habitual performance. Indeed, since the effect in the vmPFC does not appear to diminish with training, habitual behavior may come about not because the outcome value is no longer represented in the vmPFC, but rather because regions such as the DLS may come to preferentially influence behavior. That is, it appears that circuits responsible for goal-directed and habitual behavior are simultaneously engaged, but may compete for control of behavior. Habitual behavior may be produced as relative engagement of the DLS increases, even while the individual remains aware of outcome value. Indeed, rodent lesion studies show that disruption of the habit system reinstates goal-directed behavior, suggesting that goal-related representations remain intact even once the habit system has come to control behavior (Coutureau & Killcross, 2003; Yin *et al*., 2006). Similarly, habit learning need not involve a reduction in brain processing related to reward receipt. Indeed, the response in the nucleus accumbens to reward presentation remained consistent throughout the three days of training, indicating that it may process reward-related information even once control of behavior shifts toward being governed by habit rather than by the goal of obtaining the reward.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Due to our a priori hypotheses based on rodent work indicating a role of the striatum in habit learning, we have focused on our results in the DLS. However, other regions in the cortex also showed a significant increase in sensitivity to the task-relevant fractals over the course of training (Table S1). For example, regions were identified in the temporal cortex in both our between-subjects and within-subjects analyses. The inferior temporal cortex has been implicated in the formation of visuomotor associations in monkeys (Mishkin *et al*., 1984) and is connected with the caudal putamen and tail of the caudate through the "visual" corticostriatal loop (Middleton & Strick, 1996). Although our interpretation of the role of these regions in our study must remain speculative, our results point to candidate regions for further study on corticostriatal networks involved in the development of habits in humans.

Our study provides evidence that stimulus-driven, outcome insensitive habits can be shown to be present in humans following overtraining on a VI reward schedule. The finding that persistent outcome insensitive behavior can be induced even in healthy human subjects may have important implications for research into the etiology and treatment of a range of human neuropsychiatric diseases thought to involve impairments in habitual control, such as drug addiction, pathological gambling and obsessive compulsive disorder (Graybiel & Rauch, 2000; Goudriaan *et al*., 2004; Everitt & Robbins, 2005). Moreover, our finding that the development of these habits correlates with activity changes in the DLS identifies a specific neuroanatomical target for subsequent research into the neural mechanisms underlying habitual behavior in both adaptive and maladaptive contexts.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Saori Tanaka, Signe Bray, Jan Gläscher, Nicole Tetrault, and Ralph Lee for their assistance. This work was supported by grants from the Gordon and Betty Moore Foundation and the National Institute of Mental Health (RO3MH075763) to J.O.D.

Abbreviations

BOLD	Blood Oxygen Level Dependent
DLS	dorsolateral striatum
fMRI	functional magnetic resonance imaging
S-R	stimulus-response
vmPFC	ventromedial prefrontal cortex

References

Balleine BW, Dickinson A. Goal-directed instrumental action: Contingency and incentive learning and their cortical substrates. Neuropharmacology. 1998a; 37:407–419. [PubMed: 9704982]

Balleine BW, Dickinson A. The role of incentive learning in instrumental outcome revaluation by sensory-specific satiety. Anim. Learn. Behav. 1998b; 26:46–59.

Balleine BW, Ostlund SB. Still at the choice-point: action selection and initiation in instrumental conditioning. Ann N Y Acad Sci. 2007; 1104:147–171. [PubMed: 17360797]

Bayley PJ, Franscino JC, Squire LR. Robust habit learning in the absence of awareness and independent of the medial temporal lobe. Nature. 2005; 436:550–553. [PubMed: 16049487]

Colwill RC, Rescorla RA. Postconditioning devaluation of a reinforcer affects instrumental responding. J. Exp. Psychol. Anim. Behav. Proc. 1985; 11:120–132.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Coutureau E, Killcross S. Inactivation of the infralimbic prefrontal cortex reinstates goal-directed responding in overtrained rats. Behav. Brain Res. 2003; 146:167–174. [PubMed: 14643469]

Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ. Cortical substrates for exploratory decisions in humans. Nature. 2006; 441:876–879. [PubMed: 16778890]

Declaration of Helsinki. Human Experimentation: Code of ethics of the World Medical Association. Br. Med. J. 1964; 2:177. [PubMed: 14150898]

Deichmann R, Gottfried JA, Hutton C, Turner R. Optimized EPI for fMRI studies of the orbitofrontal cortex. Neuroimage. 2003; 19:430–441. [PubMed: 12814592]

Dickinson, A.; Balleine, B. The role of learning in the operation of motivational systems. In: Gallistel, CR., editor. Stevens' Handbook of Experimental Psychology: Learning, Motivation, and Emotion. Wiley and Sons; New York: 2002. p. 497-534.

Dickinson A, Balleine BW, Watt A, Gonzales F, Boakes RA. Overtraining and the motivational control of instrumental action. Anim. Learn. Behav. 1995; 22:197–206.

Dickinson A, Nicholas DJ, Adams CD. The effect of the instrumental training contingency on susceptibility to reinforcer devaluation. Q. J. Exp. Psychol. B. 1983; 35:35–51.

Everitt BJ, Robbins TW. Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. Nature Neurosci. 2005; 8:1481–1489. [PubMed: 16251991]

Foerde K, Knowlton BJ, Poldrack RA. Modulation of competing memory systems by distraction. Proc. Natl. Acad. Sci. U.S.A. 2006; 103:11778–11783. [PubMed: 16868087]

Forman SD, Cohen JD, Fitzgerald M, Eddy WF, Mintun MA, Noll DC. Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): Use of a cluster-size threshold. Magn. Reson. Med. 1995; 33:636–647. [PubMed: 7596267]

Garner DM, Olmsted MP, Bohr Y, Garfinkel PE. The eating attitudes test: psychometric features and clinical correlates. Psychol. Med. 1982; 12:871–878. [PubMed: 6961471]

Gottfried JA, O'Doherty J, Dolan RJ. Encoding predictive reward value in human amygdala and orbitofrontal cortex. Science. 2003; 301:1104–1107. [PubMed: 12934011]

Goudriaan AE, Oosterlaan J, Beurs E.d. Brink W.V.d. Pathological gambling: a comprehensive review of biobehavioral findings. Neurosci. Biobehav. Rev. 2004; 28:123–141. [PubMed: 15172761]

Graybiel AM. Habits, rituals, and the evaluative brain. Annu. Rev. Neurosci. 2008; 31:359–387. [PubMed: 18558860]

Graybiel AM, Rauch SL. Toward a neurobiology of obsessive-compulsive disorder. Neuron. 2000; 28:343–347. [PubMed: 11144344]

Hampton AN, Bossaerts P, O'Doherty JP. The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. J. Neurosci. 2006; 26:8360–8367. [PubMed: 16899731]

Killcross AS, Coutureau E. Coordination of actions and habits in the medial prefrontal cortex of rats. Cereb. Cortex. 2003; 13:400–408. [PubMed: 12631569]

Kim H, Shimojo S, O'Doherty JP. Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. PLoS Biol. 2006; 4:e233. [PubMed: 16802856]

Knowlton BJ, Mangles JA, Squire LR. A neostriatal habit learning system in humans. Science. 1996; 273:1399–1402. [PubMed: 8703077]

Maldjian JA, Laurienti PJ, Burdette JB, Kraft RA. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. Neuroimage. 2003; 19:1233–1239. [PubMed: 12880848]

Maldjian JA, Laurienti PJ, Burdette JH. Precentral gyrus discrepancy in electronic versions of the Talairach atlas. Neuroimage. 2004; 21:450–455. [PubMed: 14741682]

Middleton FA, Strick PL. The temporal lobe is a target of output from the basal ganglia. Proc. Natl. Acad. Sci. U.S.A. 1996; 93:8683–8687. [PubMed: 8710931]

Mishkin, M.; Malamut, B.; Bachevalier, J. Memories and habits: Two neural systems. In: Lynch, G.; McGaugh, JL.; Weinberger, NM., editors. Neurobiology of learning and memory. Guilford Press; New York: 1984. p. 65-77.

O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ. Dissociable roles of ventral and dorsal striatum in instrumental conditioning. Science. 2004; 304:452–454. [PubMed: 15087550]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

O'Doherty J, Rolls ET, Francis S, Bowtell R, McGlone F, Kobal G, Renner B, Ahne G. Sensory-specific satiety-related olfactory activation of the human orbitofrontal cortex. Neuroreport. 2000; 11:399–403. URLJ: www neuroreport com/. [PubMed: 10674494]

Orbitello B, Ciano R, Corsaro M, Rocco PL, Taboga C, Tonutti L, Armellini M, Balestrieri M. The EAT-26 as screening instrument for clinical nutrition unit attenders. Int. J. Obes. (Lond). 2006; 30:977–981. [PubMed: 16432540]

Parent A, Hazrati LN. Functional anatomy of the basal ganglia. I. The cortico-basal ganglia-thalamo-cortical loop. Brain Res. Brain Res. Rev. 1995; 20:91–127. [PubMed: 7711769]

Roesch MR, Olson CR. Neuronal activity related to anticipated reward in frontal cortex: does it represent value or reflect motivation? Ann. N.Y. Acad. Sci. 2007; 1121:431–446. [PubMed: 17846160]

Rolls E, Rolls B, Rowe E. Sensory-specific and motivation-specific satiety for the sight and taste of food and water in man. Physiol. Behav. 1983; 30:85–92.

Schoenbaum G, Chiba AA, Gallagher M. Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. Nature Neurosci. 1998; 1:155–159. [PubMed: 10195132]

Spooren WP, Lynd-Balta E, Mitchell S, Haber SN. Ventral pallidostriatal pathway in the monkey: evidence for modulation of basal ganglia circuits. J. Comp. Neurol. 1996; 370:295–312. [PubMed: 8799857]

Tanaka SC, Balleine BW, O'Doherty JP. Calculating consequences: brain systems that encode the causal effects of actions. J. Neurosci. 2008; 28:6750–6755. [PubMed: 18579749]

Tanaka SC, Doya K, Okada G, Ueda K, Okamoto Y, Yamawaki S. Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. Nat Neurosci. 2004; 7:887–893. [PubMed: 15235607]

Tricomi EM, Delgado MR, Fiez JA. Modulation of caudate activity by action contingency. Neuron. 2004; 41:281–292. [PubMed: 14741108]

Valentin VV, Dickinson A, O'Doherty JP. Determining the neural substrates of goal-directed learning in the human brain. J. Neurosci. 2007; 27:4019–4026. [PubMed: 17428979]

Yin HH, Knowlton BJ, Balleine BW. Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. Eur. J. Neurosci. 2004; 19:181–189. [PubMed: 14750976]

Yin HH, Knowlton BJ, Balleine BW. Inactivation of dorsolateral striatum enhances sensitivity to changes in the action-outcome contingency in instrumental conditioning. Behav. Brain Res. 2006; 166:189–196. [PubMed: 16153716]

Yin HH, Ostlund SB, Knowlton BJ, Balleine BW. The role of the dorsomedial striatum in instrumental conditioning. Eur. J. Neurosci. 2005; 22:513–523. [PubMed: 16045504]

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

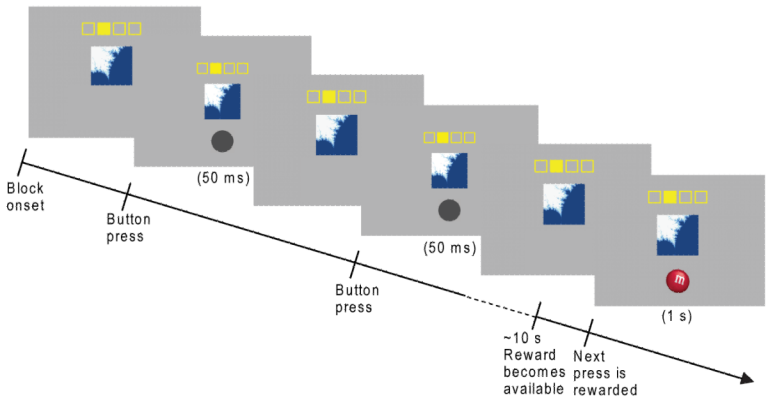

Fig. 1.

Illustration of free-operant VI-10 task. A fractal image remained present throughout the block. The filled-in yellow square indicated which button to press. Responses were self-paced. After non-rewarded responses, a dark gray circle was presented for 50 ms. A reward became available with a probability of 0.1 per second. After the subsequent button press, a picture of an M&M or Frito was shown for 1 s, indicating a food reward of the corresponding type. The two stimulus-response-outcome pairings used for each subject remained consistent throughout the experiment. A third fractal, shown with empty yellow squares, indicated a rest block.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

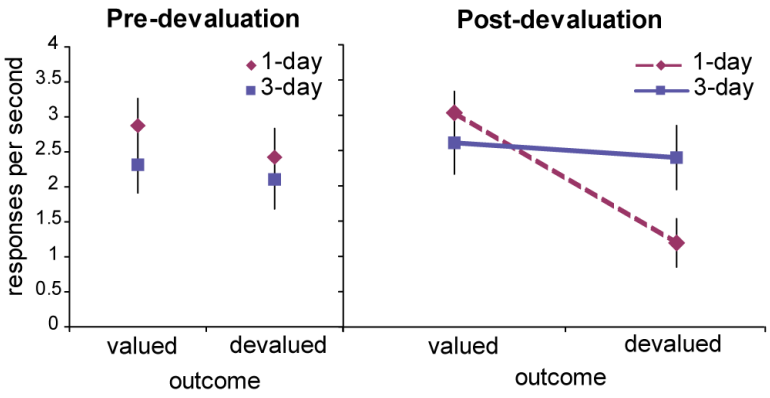

Fig. 2.

Behavioral results. During the last session of training, prior to the devaluation procedure (left), there are no significant differences in response rates between groups or when responding for the two food rewards (one which will be devalued through selective satiation and one which will not). During the extinction test following the devaluation procedure, response rates for the still-valued outcome remained high, as did response rates for the devalued outcome for the 3-day group. In contrast, response rates for the 1-day group for the devalued outcome are reduced, producing a significant training by devaluation interaction ($P < 0.05$).

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

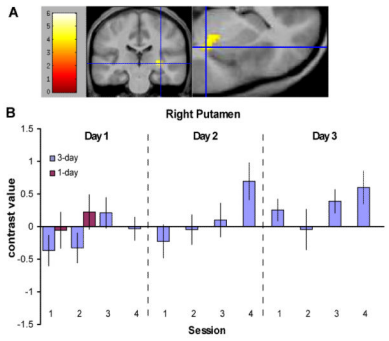

Fig. 3.

Neural correlates of habit learning, as revealed by an increasing response with training to the onset of task blocks relative to the onset of rest blocks in the 3-day group. (**A**) The right posterior putamen showed a significant increase in the [task onset — rest onset] contrast from the first two sessions to the final two sessions of training ($x = 33$, $y = -24$, $z = 0$; $P < 0.001$, $P(\text{cor}) < 0.05$). The blue crosshairs mark the voxels with the peak contrast value. (**B**) A plot of the contrast estimates for [task onset — rest onset] for each session of training is shown for the region displayed in (A). Contrast estimates for the 2 sessions from the 1-day group are also shown.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

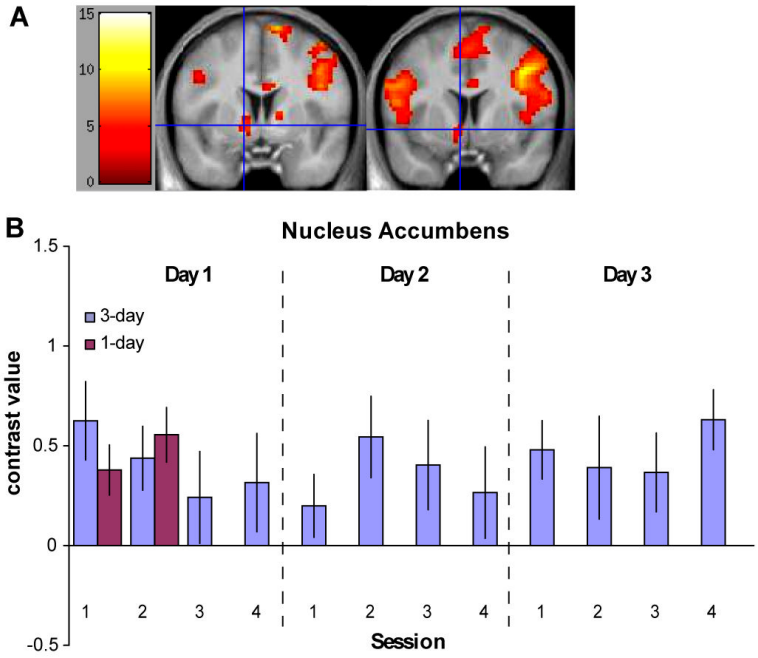

Fig. 4.

Nucleus accumbens region displaying a phasic response to reward presentation. (**A**) The nucleus accumbens displayed a significant response to reward presentation for both the 1-day group (left; $x = -12$, $y = 3$, $z = -3$; $P < 0.001$) and the 3-day group (right; $x = -6$, $y = 6$, $z = -6$; $P < 0.001$). The blue crosshairs mark the voxels with the peak contrast value. (**B**) A plot of the contrast estimates for the reward contrast for each training session is shown for the regions displayed in (A). The contrast magnitude remained consistent over training.

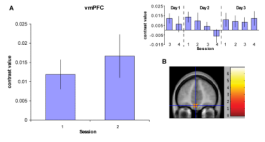

Fig. 5.

Neural correlates of goal representation, as revealed by increased activation as anticipation of an upcoming reward increases. (**A**) A plot of the contrast estimates for the region in the vmPFC showing an effect of the “ramp” modulator for the first two sessions of training for all subjects is shown. This modulator indicates the number of seconds since the previous reward presentation or start of the block during performance of task blocks. The contrast estimates for each of the remaining training sessions for the 3-day group is also shown (top right). There is no clear pattern of change in this contrast over the course of training. (**B**) The vmPFC region showing a significant effect of the ramp modulator in all subjects over the first 2 sessions of training is shown ($x = -3$, $y = 45$, $z = -18$; $P < 0.001$). The blue crosshair marks the voxel with the peak contrast value.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript