Dissociable Roles of Ventral and Dorsal Striatum in Instrumental Conditioning

John O’Doherty,1* Peter Dayan,2 Johannes Schultz,1 Ralf Deichmann,1 Karl Friston,1 Raymond J. Dolan1

Instrumental conditioning studies how animals and humans choose actions appropriate to the affective structure of an environment. According to recent reinforcement learning models, two distinct components are involved: a “critic,” which learns to predict future reward, and an “actor,” which maintains information about the rewarding outcomes of actions to enable better ones to be chosen more frequently. We scanned human participants with functional magnetic resonance imaging while they engaged in instrumental conditioning. Our results suggest partly dissociable contributions of the ventral and dorsal striatum, with the former corresponding to the critic and the latter corresponding to the actor.

The ability to orient toward specific goals in the environment and control actions flexibly in pursuit of those goals is a hallmark of adaptive behavior. Instrumental conditioning, the most basic form of such behavior, allows an organism to learn contingencies between its own responses and rewarding or punishing outcomes (1–3). Models of reinforcement learning, such as the actor-critic (4) or advantage learning model (5), provide a two-process account of instrumental conditioning. One component, the critic, uses a temporal difference prediction error signal to update successive predictions of future reward associated with being at a state of the external and internal environment (determined by the arrangement of stimuli). The other component, the actor, uses a similar signal to modify stimulus-response or stimulus-response-reward associations in the form of a policy, so that actions associated with greater long-term reward are chosen more frequently on subsequent trials (8–11).

A putative neuronal correlate of these temporal difference prediction error signals is the phasic activity of dopamine neurons (12–14), which send prominent projections to the ventral and dorsal striatum. Lesion and human imaging studies suggest that the ventral and dorsal striatum may have distinct functions. The former is implicated in reward and motivation (15). The latter is implicated in motor and cognitive control (16–19), specifically the learning of stimulus-response associations. On the basis of these findings, a putative neural substrate for reinforcement learning has been proposed (20), according to which dopaminergic projections to ventral striatum might be involved in reward prediction, corresponding primarily to the critic component of instrumental learning, whereas dopaminergic projections to dorsal striatum might be involved in the modulation of stimulus-response or stimulus-response-reward associations, corresponding to the instrumental actor.

We analyzed functional magnetic resonance imaging (fMRI) data from human participants performing an instrumental conditioning task. We used a reinforcement learning model called advantage learning (21) to calculate a reward prediction error signal and tested for correlations between that signal and evoked neural activity in the striatum. To dissociate stimulus-response learning from value prediction learning itself, we used a yoked Pavlovian conditioning task as a control condition. This task involves the same value predictions (critic), without action selections (actor). If the ventral striatum corresponds to the critic, then this region should show prediction error activity during both the instrumental and Pavlovian conditioning tasks. If the dorsal striatum corresponds to the actor, then we would expect it to manifest stronger prediction error–related activity during instrumental than during Pavlovian conditioning.

The instrumental task was composed of two trial types: reward and neutral. In the reward trials, participants had to choose between one of two stimuli: one associated with a high probability of obtaining a juice reward (on 60% of occasions) and the other with a low probability of obtaining a juice reward (on 30% of occasions). In neutral trials, participants had to choose between two other stimuli associated with either a high (60%) or low (30%) probability of obtaining an affectively neutral solution. The Pavlovian task was identical to the instrumental task (with both reward and neutral trials), except that the computer made the selection and the participant’s task was to indicate which stimulus had been chosen by the computer (Fig. 1A).

Participants rated the fruit juice as significantly more pleasant than the control tasteless solution in both the instrumental and Pavlovian conditioning tasks (P < 0.001; Fig. 1B). In the reward trials of the instrumental task, participants chose the high-probability action significantly more frequently than the low-probability action, but they showed no preference for the high-probability action in the neutral trials (Fig. 1C). There was evidence of “response matching” in the instrumental task (22) in that the ratio of responses made to the high-probability and low-probability stimuli was 1.92:1 during reward trials, a value very close to the actual 2:1 ratio of reward probabilities associated with the two stimuli.

To obtain a behavioral measure of learning in the Pavlovian conditioning task, we tested for differences in reaction times between responses in the reward and neutral trials (pooling over responses to high- and low-probability stimuli) between early and late phases of the session. Participants were faster to respond during the reward trials than during the neutral trials by the second block of trials (Fig. 1D).

References

5. We thank D. Eliezer and T. McGraw for helpful discussions, K. D’Amico and L. Tong for help with data collection at the Advanced Photon Source, and D. Landry for help with the surface plasmon resonance experiment. This work was supported by P01AG08702, AG17636, AG17490, and R01DA07145 and in part by Speaker’s Fund for Biomedical Research and NIH 1K07AG00959. H.W. is a Pew Scholar in biomedical sciences and a Rita Allen Scholar.

*To whom correspondence should be addressed. E-mail: j.odoherty@filion.ucl.ac.uk

1Wellcome Department of Imaging Neuroscience, Institute of Neurology, Gatsby Computational Neuroscience Unit, University College London, London WC1N 3BG, UK.
trials than neutral trials (approaching significance at P < 0.05). We first replicated previous findings of reward prediction error activity in the ventral striatum (ventral putamen) during Pavlovian conditioning (Fig. 2A) (23, 24). This extends the previous results, because here we compared prediction error responses between high- and low-valence gustatory stimuli, both of which involve sensory stimulation in the mouth and orofacial movement. Consequently, we now control for somatomotor effects and demonstrate that prediction error activity in the ventral striatum is specific to an affectively significant stimulus.

Next, we analyzed the instrumental conditioning task. Figure 2B shows that the blood oxygen level–dependent (BOLD) signal in a part of the ventral striatum, the nucleus accumbens, is correlated with the prediction error signal during the instrumental task (P < 0.001), consistent with our hypothesis that, because of its association with the critic, the ventral striatum is recruited during instrumental as well as Pavlovian conditioning. Figure 2C shows the results of a direct test of common activity during both forms of conditioning, confirming the involvement of the nucleus accumbens and ventral putamen, which are both parts of the ventral striatum (P < 0.001).

We also tested for significant prediction error activity in the dorsal striatum during instrumental conditioning. The BOLD signal in the anterior caudate nucleus, a region of the dorsal striatum, was significantly correlated with the instrumental prediction error signal at P < 0.001 (Fig. 3A). Significant effects were not found in this area in the Pavlovian conditioning task (even at P < 0.01). By subtracting prediction error responses expressed during Pavlovian conditioning from those expressed during instrumental conditioning, we showed that predic-

**Fig. 1.** (A) Illustration of instrumental task. Participant chose one of two fractals, which on each trial were randomly assigned to the left or right of the fixation cross. After the choice, the chosen fractal was illuminated, and 2000 ms later the outcome occurred. After another 3000 ms, the next trial was triggered. (B) Pleasantness ratings for the fruit juice and control tasteless solutions. Ratings were taken before and after the instrumental and Pavlovian conditioning sessions (+10, very pleasant; 0, neutral; −10, very unpleasant). Participants found the fruit juice to be significantly more pleasant than the control tasteless solution (P < 0.001). (C) Choices of high- versus low-probability actions in the instrumental task. Plot shows total number of choices of the high-probability (HP) and low-probability (LP) actions averaged across participants in both the reward and neutral trials of the instrumental conditioning task. Participants chose the high-probability action significantly more often than the low-probability action in reward trials (P < 0.05). (D) Reaction times during the Pavlovian conditioning task. Differences in reaction times are shown plotted between the reward and neutral trials during the Pavlovian conditioning task. In the second phase of the experiment, participants were faster to respond during reward trials than neutral trials (approaching significance at P = 0.054). This provides a behavioral measure of learning, providing some evidence that participants did acquire the Pavlovian associations. Error bars show mean ± SEM.

**Fig. 2.** Ventral striatum correlating with prediction error signal during Pavlovian and instrumental conditioning. (A) Reward prediction error responses in bilateral ventral striatum (ventral putamen) during Pavlovian conditioning in reward compared to neutral trials (left hemisphere coordinates: −26, 8, −4 mm; peak z-score = 3.98; right hemisphere coordinates: 26, 6, −8 mm; z = 4.167). Effects significant at P < 0.001 are shown in yellow, and effects significant at P < 0.01 are shown in red to illustrate the full extent of the activation. R, right. (B) Reward prediction error responses in ventral striatum (nucleus accumbens) during instrumental conditioning (right hemisphere coordinates: 6, 14, −2 mm; z = 3.43). (C) Results are shown for the conjunction of the prediction error signal for both types of conditioning. Significant effects were found in bilateral ventral striatum [in the bilateral ventral putamen (left hemisphere coordinates: −28, 8, −6 mm; z = 3.73; right hemisphere coordinates: 20, 12, −8 mm; z = 3.54) and in the right nucleus accumbens (14, 10, −10 mm; z = 3.21)] at P < 0.001. Images in (A), (B), and the left and middle panels of (C) show coronal slices through different sections of ventral striatum (at y = 8 mm, y = 14 mm, y = 8 mm, y = 10 mm, respectively). A plot of the contrast estimates is also shown (bar chart, right) for the peak voxel in the conjunction analysis with prediction error (PE) effects at the time of presentation of the cue or conditioned stimulus (cs) and at the time of presentation of the reward or unconditioned stimulus (ucs), plotted separately for each type of conditioning.
nucleus were significantly enhanced in in-
duction error responses in the left caudate 
during instrumental 
discrimination error signal 
correlating with pre-
icephalic complex of the 
neural activity in the 
neural activity in the 
neutralization of the predi-
ing signals used in instrumental 
stimulus-prediction error signals used in instrumental 
prediction error signals used in instrumental 
conditioning than in Pavlovian 
conditioning (P < 0.001; −6, 22, 
2; z = 3.78) (left). A plot of the contrast estimates is also shown (right) for the peak voxel.

dition error responses in the left caudate 
nucleus were significantly enhanced in in-
strumental conditioning at P < 0.001 (Fig. 
3B). The finding of enhanced temporal-
difference prediction error–related responses in 
dorsal striatum during instrumental condi-
tioning compared with Pavlovian condition-
ing is consistent with our hypothesis that this 
region plays a central role in implementing the 
instrumental actor.

Activity in both the ventral and dorsal 
striatum during instrumental-styled tasks has 
been shown previously in which a response is 
required to obtain an outcome (25, 26) but in 
which, notably, there is no explicit choice 
between different allowed responses. Our 
task embodies the essence of instrumental 
conditioning in that the participants had to 
choose between two options (overall favoring 
the high-probability action), which required 
their active engagement. An fMRI study of 
gambling in which participants were engaged 
in decision making has also reported activity 
in the dorsal striatum (27). Our results dif-
f erentiate ventral and dorsal striatum ac-

ting to their relative contributions to 
stimulus-reward and stimulus-response (or stimulus-response-reward) learning. Anato-
mical distinctions, such as those made be-

tween matrisomes and striasomes within the 
dorsal striatum (28, 29), have also been im-
plicated in the implementation of critic and 
actor learning (9, 30) on the basis of their 
differential control over and innervation by 
dopamine. However, these lie at a finer spa-
tial scale than is accessible with the resolu-
tion of our neuroimaging technique.

Advantage learning, which underlies the 
prediction error signals used in instrumental 
conditioning, has been suggested as a bridge 
between goal-directed action selection, in 
which actions are chosen with reference to an 
explicit representation of the incentive value 
of the outcome or goal state, and habitual 
action selection, in which actions are elicited 
by the presentation of a specific stimulus, 
without incorporating a representation of the 
outcome itself (3, 31). Although we did not 
directly test it here, goal-directed forms of 
instrumental learning may rely on structures 
in the prefrontal cortex (4, 32).

Reinforcement learning links psychological 
ideas of stimulus-reward and stimulus-
response-reward learning to computational 
and engineering ideas about adaptive optimal 
control (6) and a putative dopaminergic sub-
strate (20, 33). The present study on instru-
mental conditioning, together with previous 

Fig. 3. Dorsal striatum 

correlating with predi-

cion error signal 
during instrumental 
conditioning. (A) Re-
sults depict the corre-
lation of the predic-
tion error signal with 
nuclear activity in the 
dorsal striatum for the 
instrumental task (left) and the Pavlov-
ian task (right). Signif-
ificant activations were 
found in the left ante-
rior caudate nucleus 
(−8, 22, 0; z = 3.84).
No significant effects were 
observed in the Pavlovian task at P < 
0.001 or even P < 
0.01. R, right. (B) Area 
of the dorsal striatum 
(anterior caudate nu-
cleus) showing signifi-
cantly greater predic-
tion error responses in instrumental conditioning than in Pavlovian conditioning (P < 0.001; −6, 22, 
2; z = 3.78) (left). A plot of the contrast estimates is also shown (right) for the peak voxel.

dicated control over and innervation by 
dopamine. However, these lie at a finer spa-
tial scale than is accessible with the resolu-
tion of our neuroimaging technique.

Advantage learning, which underlies the 
prediction error signals used in instrumental 
conditioning, has been suggested as a bridge 
between goal-directed action selection, in 
which actions are chosen with reference to an 
explicit representation of the incentive value 
of the outcome or goal state, and habitual 
action selection, in which actions are elicited 
by the presentation of a specific stimulus, 
without incorporating a representation of the 
outcome itself (3, 31). Although we did not 
directly test it here, goal-directed forms of 
instrumental learning may rely on structures 
in the prefrontal cortex (4, 32).

Reinforcement learning links psychological 
ideas of stimulus-reward and stimulus-
response-reward learning to computational 
and engineering ideas about adaptive optimal 
control (6) and a putative dopaminergic sub-
strate (20, 33). The present study on instru-
mental conditioning, together with previous 

References and Notes

1. B. F. Skinner, The Behavior of Organisms (Appleton-
Century-Crofts, New York, 1938).
2. E. L. Thordike, Animal Intelligence: Experimental 
Studies (Macmillan, New York, 1911).
3. N. J. Mackintosh, Conditioning and Associative Learn-
4. B. W. Balleine, A. Dickinson, Neuropsychopharmacology 37, 
6. R. S. Sutton, A. G. Barto, Reinforcement Learning (MIT 
7. L. C. Baird, Advantage Updating (Rep. No. WL-TR-93- 
1146, Wright Patterson Air Force Base, Dayton, OH, 
1993).
8. A. G. Barto, in Models of Information Processing in the 
9. J. C. Houk, J. L. Adams, A. G. Barto, in Models of 
Information Processing in the Basal Ganglia, J. C. Houk, J. L. Davis, D. G. Beiser, Eds. (MIT Press, Cam-
11. D. Joel, Y. Niv, E. Ruppin, Neuronal Networks 15, 535 
(2002).
(1994).
15. R. N. Cardinal, J. A. Parkinson, J. Hall, B. J. Everitt, 
Neurosci. 9, 357 (1986).
25, 563 (2002).
19. I. Toni, R. E. Passingham, Exp. Brain Res. 127, 19 
(1999).
16, 1936 (1996). Also see discussion in (17).
21. Materials and methods are available as supporting 
material on Science Online.
24. S. M. McClure, G. S. Berns, P. R. Montague, Neuron 38, 
25. G. Pagnoni, C. F. Zink, P. R. Montague, G. S. Berns, 
26. B. Knutson, G. W. Fong, C. M. Adams, J. L. Varner, D.
27. M. R. Delgado, H. M. Locke, V. A. Stenger, J. A. Fiez, 
30. J. Brown, D. Bullock, S. Grossberg, J. Neurosci. 19, 
10502 (1999).
31. P. Dayan, in Advances in Neural Information Processing 
Systems 14, T. G. Dietterich, S. Becker, Z. Ghahra-
11–18.
32. S. Killcross, E. Coutureau, Cereb. Cortex 13, 400 
(2003).
33. W. Schultz, P. Dayan, P. R. Montague, Science 275, 
1593 (1997).
34. We thank B. Seymour and R. Passingham for helpful 
discussions. Supported by Wellcome Programme 
grants (R.J.D. and K.F.). P.D. is supported by the 
Gatsby Charitable Foundation.

Supporting Online Material 
www.sciencemag.org/cgi/content/full/304/5669/452/ 
DC1 
Materials and Methods 
Figs. S1 to S4 
References and Notes 

3 December 2003; accepted 15 March 2004