Habitual control of goal selection in humans

Fiery Cushman[1] and Adam Morris[2]

[1]*Department of Psychology, Harvard University*

[2]*Department of Cognitive, Linguistic and Psychological Science, Brown University*

The distinction between habitual and goal-directed action is fundamental to behavioral research[1-3]. Habits form as stimulus-response pairings are "stamped in" following reward. They enable computationally efficient decision making, but at the cost of behavioral flexibility. In contrast, goal-directed behavior requires planning over a causal model. This enables more flexible decision-making, but at a potentially severe computational cost. Exhaustive search over candidate goals becomes prohibitive as the complexity of the model grows. Thus, a key requirement for goal-directed action is to efficiently select candidate goals with a high likelihood of reward. Here, we provide evidence for a potential solution implemented by humans: Habitual control over the process of goal selection. Although many existing treatments of the distinction between habitual and goal-directed action emphasize their competition over behavioral control[4], our results illustrate a codependence between the systems. The role of habitual control in goal-directed action explains diverse phenomena including the automatic selection of goals under contextual cuing[5,6] and the basis of practice effects in cognitive skills[7,8].

Our approach depends on a formalization of goal-directed and habitual behavior derived from the reinforcement learning (RL) framework[9]. Model-based RL maintains an explicit causal model of the world and uses it to choose actions by assessing their likely consequences, thus enabling goal-directed planning. In contrast, model-free RL does not maintain an explicit causal model, and therefore does not allow planning. Rather, like Thorndike's law of effect[2], it assigns value to candidate actions based on their context-dependent history of reward. The resulting stimulus-response habits are globally adaptive, but may exhibit local irrationality[10,11].

RL models capture several core features of learning and choice in humans[1,10]. Elements of model-free RL, including prediction-error updating and temporal difference learning, are implemented in the midbrain dopamine system[12-14]. Human behavior also relies extensively on model-based planning towards goals, often arranged hierarchically (plug in the machine to grind the beans to make the coffee, etc.)[3,15,16]. Once selected, goals can dramatically reduce the computation necessary for model-based action selection[17,18]. The resulting policies are then available for reuse and recombination, further reducing computational demands[19]. First, however, an appropriate goal must be selected. Full model-based evaluation of candidate goals often imposes a prohibitive computational cost. In principle, given the superordinate goal of making coffee, an infinite number of subordinate goals might be entertained and evaluated. How do we efficiently alight upon the next relevant goal: ground beans?

One potential solution is to allow model-free control over goal selection. In other words, the goal of grinding beans might be "stamped in" due to the history of reward associated with this goal in past coffee-making episodes. Subsequent planning to achieve the selected goal could then proceed in a model-based fashion, or by other methods. This enables computationally tractable goal selection while maintaining the potential for flexible planning towards the selected goal.

Colloquially, this proposal captures the notion of a "habit of thought": Model-free control can contribute to the effective deployment of model-based cognitive routines, which themselves transcend learned stimulus-response pairings[16]. Consistent with this proposal, recent research emphasizes the pervasive role of model-free control in related elements of higher-level cognition[20,21], including the gating of working memory[22] and the construction of hierarchical task representations[19]. Such proposals offer an appealing functional explanation for the neuronal connections between striatum and frontal cortex[23]. The possibility of habitual control over goal selection also accords with several formal models of RL[24] and human cognition[3].

In order to test this possibility we adapted a multistep choice paradigm from prior research[10]. The original paradigm behaviorally dissociates the influence of habitual (model-free) and goal-directed (model-based) control on choice, and is well-validated[4,25-27]. Our modification allows us to index not only model-free value assignment to actions (as in the original task) but also model-free value assignment to goals, which may be subsequently pursued via model-based planning (Figure 1a). At Stage 1 of each trial participants make a choice between two actions drawn from the set (1, 2, 3, 4). These choices stochastically transition to Stage 2 states from the set (blue, red, green). Finally, Stage 2 states deterministically transition to three unique reward distributions. The rewards change gradually over the course of the experiment. Thus, participants are motivated to choose Stage 1 options that maximize the likelihood of transitioning to the current reward-maximizing final state. Participants received with detailed instructions and practice trials, including information about the stochastic transitions between Stage 1 and Stage 2. For task details, see Supplementary Information.

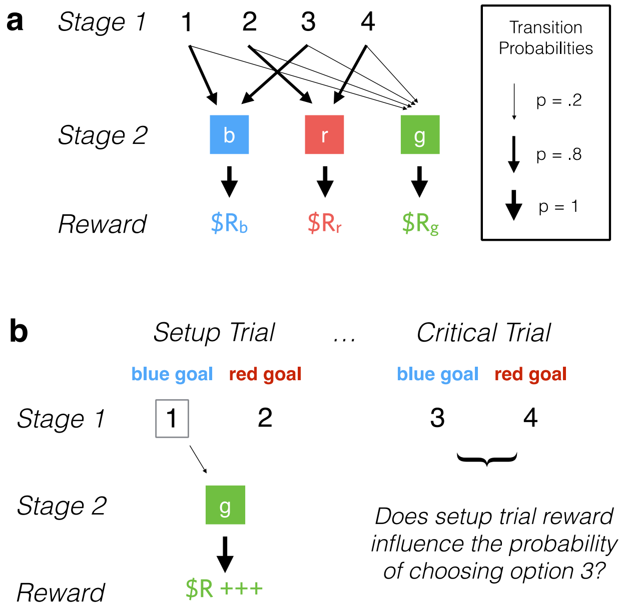

Figure 1: Experiment 1 structure and logic. **a**, In Experiment 1 participants performed a two-stage Markov decision task. They were presented with two possible Stage 1 actions drawn from a set of four. These transitioned with variable probabilities to a set of Stage 2 actions, which then transitioned deterministically to a set of drifting reward distributions. **b**, The logic of the experiment depends on a subset of critical trials. For instance, participants might be presented with the choice set (1,2) in a setup trial. Upon selecting action 1, they experience a low-probability transition to the green state and followed by a large reward. A model-free influence on goal selection uniquely predicts an increase in the selection of action 3 on the subsequent critical trial, because actions 1 and 3 share the common goal state of blue.

Our analysis depends on a critical subset of trials (Figure 1b). For example, a participant is presented with the choice set (1, 2) at Stage 1 and chooses action 1. Because 1 typically leads to the blue state, we assume that this participant's goal was to transition to blue. On our critical trials, however, they experience a low-probability transition to the green state, and then experience a very large reward. A model-based system would discard this information because transitions to the green state are equally likely from all Stage 1 options. This renders forward planning toward green irrelevant. In contrast, model-free value update would increase the likelihood of selecting 1 on subsequent trials due to the positive reward history[10]. Our interest, however, is in the model-free assignment of value to a goal; in this case, the goal of transitioning to blue. If the experience of reward increases the likelihood of selecting blue as a goal, then participants should exhibit a greater likelihood of choosing 3 on the subsequent trial (when paired with either 2 or 4). Conversely, the experience of punishment should decrease the likelihood of choosing 3. This influence of the reinforcement history of choosing 1 on the subsequent choice of 3 cannot be explained by model-free update of a value to the specific action (choosing 1); rather, it depends on the assignment of value to their shared goal (getting to blue).

We assessed trials of this type by comparing instances when the participant experienced reward vs. punishment following low-probability transition to the green state in a setup trial. The mean proportion of trials on which participants selected the congruent-goal action following positive reward (89%) was significantly greater than the proportion following negative reward (69%) $t(134)=-12.5$, $p<.0001$ (Figure 2a). Additional supporting analyses for all experiments are presented in Supplementary Information.

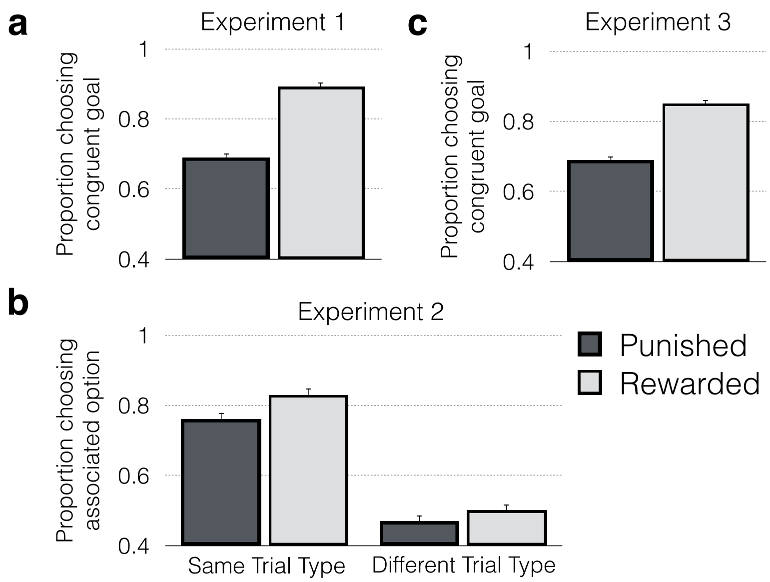

Figure 2: Results. Bars represent the proportion of trials on which participants chose the congruent-goal action, averaged across participants. Whiskers indicate the standard error of the mean of these proportions across participants. **a-c** show results from Experiments 1-3, respectively. **b**, same trial type bars show the effect of reward experienced on trials of the same type (shape vs. color) while different trial type bars show the effect of reward experienced on trials of the opposite type.

The evidence from Experiment 1 is ambiguous between two interpretations. It may be that people assign value to the selection of a goal (e.g., “choose blue”), or it may be that people assign value directly to the congruent-goal Stage 1 action (e.g., “choose option 3”). Experiment 2 was designed to disambiguate these possibilities (Figure 3). Specifically, Stage 2 states were arranged in a 2×2 design crossing color (red versus blue) and shape (circle versus square), with a fifth state that differed on both dimensions (a green triangle). Each trial was defined as a “color trial” or “shape trial”, with the trial type dictating the deterministic transitions to drifting rewards. Participants were cued to trial type at the beginning of each trial. Thus, three reward distributions were accessible on color trials, while three independent reward distributions were accessible on shape trials. This made color goals relevant only to color trials, and shape goals relevant only to shape trials. If model free value is assigned to goals, it should only influence choice on subsequent trials of the same type (i.e., "color trial" vs. "shape trial"). Consistent with this prediction, we replicated our result from Experiment 1 for same-type trials: After a low-probability transition to the green triangle state, participants were more likely to choose the congruent-goal Stage 1 action on a subsequent same-type trial following reward (83%) compared with punishment (76%) $t(302)=-4.82$, $p<.001$. On different-type trials, however, there was no significant difference (positive: 50%; negative: 47%) $t(282)=-.94$, $p=.35$ (Figure 2b).

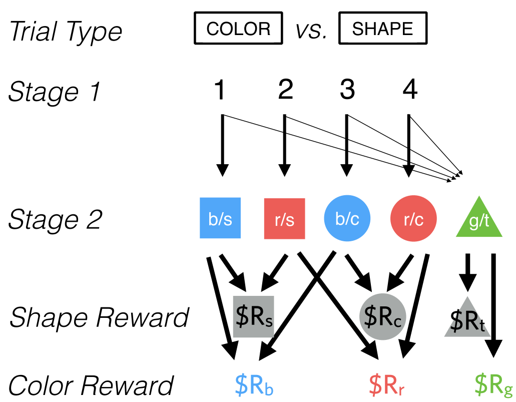

Figure 3: Experiment 2 structure. In Experiment 2, Stage 2 states varied along two orthogonal dimensions: shape and color. On each trial the participant was cued that rewards would be determined by shape, or else by color, for that trial. We predicted that rewards obtained following low-probability transitions to the green state would only influence subsequent choice on critical trials of the same trial type (shape vs. color). This is because the goals selected in each of two trials can only match when their trial types are identical.

The results of Experiments 1 and 2 are consistent with model-free control over goal selection in which the habitually selected goal then participates in a process of forward planning over a causal model of the task's transition structure. They are also consistent, however, with some hierarchical reinforcement learning models that do not invoke true model-based planning[16,28]. These models assume that "goal states" establish internally represented contexts that bias model-free stimulus-response associations. We designed Experiment 3 to test whether the goals selected in our paradigm could be flexibly integrated with knowledge of independent state transitions, a hallmark of true planning.

Participants were first trained on a deterministic set of transitions between four Stage 0 options (A, B, C, D) and the same four Stage 1 options used in Experiment 1 (Figure 4a). During this training phase the Stage 1 options comprised the terminal states. Then, participants trained on and performed the same task used in Experiment 1, without any involvement of Stage 0 choices. Finally, we tested each participant in a set of critical trials—those following the setup of a low-probability transition followed by a congruent-goal choice—but presented participants on critical trials with a pair of Stage 0 choices in place of the ordinary Stage 1 choices (Figure 4b). In order to integrate information about a desired goal with the set of Stage 0 choices, participants were required to engage in forward planning over the learned transition structure between Stage 0 and Stage 1. Here, again, we found that participants were significantly more likely to choose the congruent-goal action following positive reward (85%) than following negative reward (69%) $t(172)=-9.17$, $p<.0001$ (Figure 2c). These results also speak against an alternative interpretation of Experiments 1 and 2 according to which statistical structure of state transitions could support associative pairings between congruent-goal Stage 1 options, and thus associative transfer of reward values[29]. In Experiment 3 there is no such basis for statistical association between Stage 0 actions analyzed on the critical trial and the Stage 1 action rewarded on the setup trial.

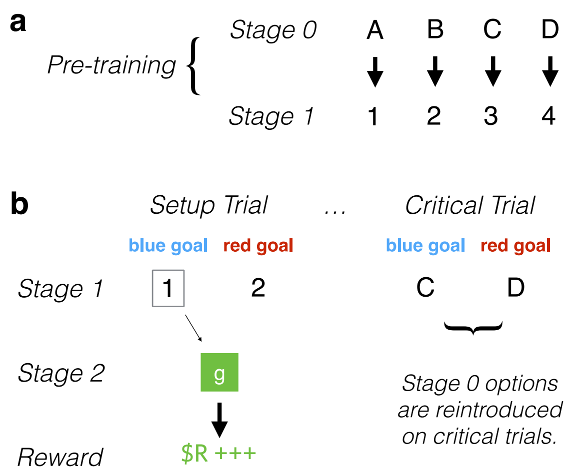

Figure 4: Experiment 1 structure and logic. Experiment 3 was modeled on the design of Experiment 1, except that **a**, participants performed a pre-training in which they learned deterministic transitions between Stage 0 and Stage 1 choices, and **b**, on critical trials the Stage 0 choices were selectively reintroduced. Thus, in order to make successful choices on critical trials, participants were required to choose a Stage 0 option that would lead to their preferred Stage 1 state.

In sum, we find that goal selection is partially determined by model-free value representations derived from reward history. These goals are subsequently used during model-based planning over an internally represented causal model of the task structure. Thus, while our proposal relies upon the conceptual distinction between habitual (model-free) and goal-directed (model-based) behavioral control, it also demonstrates an area of mutual dependence between these mechanisms.

This proposed integration captures several phenomena that blend canonical features of habits and goals. Contextual cues can trigger goal pursuit outside of conscious awareness[5], consistent with the operation of stimulus-response habits in the process of goal selection. In cases of “utilization behavior” among individuals with insult to prefrontal cortex, goal-directed behavior may be intrusive or inappropriately invoked based on contextual cues[6]. Among neurotypical individuals, “functional fixedness” describes the tendency to consider a limited set of candidate means-end relationships based on past experience with a tool[8]. Finally, it is observed in educational settings that the execution of controlled cognitive processes improves with practice—in other words, that learning complex tasks requires the incremental acquisition of appropriate habits of thought[7]. It is widely recognized that humans' representations of complex tasks are organized hierarchically into goals and subgoals[15,16]. Task proficiency may depend partially upon the gradual acquisition of habitual subgoal selection given the contextual state of a superordinate goal.

Methods

Task Our task was closely modeled on a multistep choice paradigm used in prior research to dissociate model-free and model-based contributions to human behavioral control[10]. Here we present details of the task in its general from; specific versions implemented in each experiment are described in Supplementary Information. At each stage of the multichoice procedure participants had four seconds to select one of the two available options by clicking on it with their mouse. The reward distributions following Stage 2 actions were initialized uniformly at random on a range of -4 points to +5 points, and varied according to a bounded Gaussian random walk for the remainder of the experiment. After each round, the drift was sampled from a normal distribution (μ=0, σ=1.8), rounded to the nearest integer, and added to the current reward level. When this procedure selected a reward level outside the bounds of [-4,5], the amount of excess was subtracted from the bound to obtain the new reward value.

Procedure Across our three experiments we recruited 960 participants through the Amazon Mechanical Turk online labor marketplace. Participants provided informed consent and completed an extensive instruction and training procedure. During this procedure we described successively more complex elements of our task, interspersing three sets of 25 practice trials each that allowed participants to consolidate their incremental knowledge of the task. Performance on these practice trials was not rewarded. Following instructions and training participants completed 175 rewarded trials, earning 1 cent in bonus pay for each net point accumulated.

Analysis Participants were excluded from analysis if they failed to respond within the allotted 4 second window on 50 or more rewarded trails. Following prior research[30] we also excluded participants who did not meet a minimum threshold of learning. We ran a Monte Carlo simulation of 10,000 agents performing the task randomly and determined the 95% percentile of performance (as indexed by their final scores), and excluded participants whose scores fell below that threshold.

Our analysis was restricted to a critical subset of "congruent goal" trials. We defined these according to two characteristics: they immediately followed a low-probability transition to green, and they did not present participants with the Stage 1 choice that they had chosen on the previous trial. We defined the "model-free goal value" as the reward obtained on the previous trial; that is, the most recent reward that immediately followed selection of the relevant goal. Although formal approaches to model-free reinforcement learning (e.g. Q learning) typically estimate value according to a geometrically-weighted sum of all past rewards[9], past experimental research indicates more robust statistical estimates of model-free value assignment under the simplifying assumption that the most recent reward experience dominates value representation. This estimation technique has been used in past studies of stochastic sequential decision-making paradigms[4,10]. In order to further validate this analytic approach, we show in Supplementary Information that it successfully recovers evidence for model-free value assignment to goal selection from the data generated by a formal computational model of our hypothesized mechanism.

As a course-grained analysis, for each participant we computed the proportion of congruent-goal trials on which they maintained their choice of goal, comparing trials on which model-free goal value was less than 0 (i.e. goal selection was followed by punishment) to trials on which the mean choice when it was greater than 0 (i.e. goal selection was followed by reward). We then compared these proportions across participants using a repeated measures t-test.

As a more granular test, we regressed choice on the model-free goal value using a logistic mixed-effects model, estimating both random intercepts and random slopes at the subject level. Our results, presented in Supplementary Information, confirmed those obtained by course-grained analyses. To definitively rule out a potential confounding influence from pure model-based or model-free action valuation, we estimated a second mixed-effects model with approximate model-based and model-free action values as additional regressors. Consistent with our analytic approach, the model-based value of an action with a certain color goal was approximated by the last reward that the subject received from that color. The model-free value of an action was approximated by the reward received the last time the subject selected that action. In each case, convergence of our statistical model depended upon discounting reward values; we implemented a discounting parameter of $\gamma = .85$ per trial.

We used the Wald test to derive the significance of the model-free goal regressor in the mixed-effects models. We also estimated null models (the full model with the model-free goal regressor removed), and performed both likelihood ratio tests and parametric bootstrap analyses to assess whether the model-free goal regressor increased the model's likelihood enough to justify inclusion.

All mixed-effects analyses were conducted in R[31], making use of the lme4 linear mixed effects package[32].

Model To validate our analytic approach we specified a computational model of learning and choice embodying habitual control of goal selection. This model included traditional model-based and model-free control, along with a mechanism for model-free value update and control over goal selection. We used this computational model to generate simulated data for Experiment 2, and showed that our observed results were obtained if and only if the computational model included model-free goal selection. The results of this simulation are presented in Supplementary Information.

The game was implemented as a Markov decision process with six states, the initial Stage 1 state and five Stage 2 states. The Stage 1 state had four possible actions (i.e. the four numbers), only two of which were available on any given trial. The Stage 2 states had only one possible action (i.e. clicking on the object), which led to a reward. The rewards were randomly generated for each agent by the same process as in the behavioral tasks.

The agents had three learning mechanisms. First, a model-free reinforcement learning mechanism employed the SARSA algorithm with eligibility traces[9]. Agents estimated a model-free value of the state-action pair (s,a), denoted $MFV_{(s,a)}$. In Stage 1, agents chose an action a and transitioned to state s. The value update for $MFV_{(1,a)}$ occurs by temporal difference learning with learning rate α:

$$MFV_{(1,a)} \leftarrow MFV_{(1,a)} + \alpha(MFV_{(s\prime,a\prime)} - MFV_{(1,a)})$$

In stage 2, agents chose the only available action a' (i.e. clicking on the object) and received reward r. Again, value update is given by temporal difference learning:

$$MFV_{(s,a\prime)} \leftarrow MFV_{(s,a\prime)} + \alpha(r - MFV_{(s,a\prime)})$$

In keeping with prior computational models of stochastic sequential decision-making paradigms[4,10], we also implemented an update of Stage 1 value representations following reward by applying an eligibility trace λ:

$$MFV_{(1,a)} \leftarrow MFV_{(1,a)} + \lambda\alpha(r - MFV_{(1,a)})$$

Second, agents' model-based learning mechanism implemented a basic forward planning algorithm. Agents maintained a model-based value of each state-action pair, denoted $MBV_{(s,a)}$. We assumed knowledge of the trial-type-dependent reward distributions on the part of the model-based controller. Thus, we separately indexed

Stage 2 states according to the relevant trial type. To calculate the model-based value of each action from state 1, agents estimated the transition probability from *a* to *s*, denoted $T_{(a,s)}$, by counting the number of observed transitions from *a* to *s* and normalizing the counters. (Counters were initialized to 1). Then:

$$MBV_{(1,a)} = \sum_{k=2}^{6} T_{(a,k)} * MBV_{(k,a')}$$

After transitioning to state *s*, performing the only available action *a'*, and receiving reward *r*, the model-based update was:

$$MBV_{(s,a')} \leftarrow MBV_{(s,a')} + \alpha(r - MBV_{(s,a')})$$

The third learning mechanism was our proposed mechanism, implementing model-free value update and control over goal selection. After a trial with chosen action *a* and received reward *r*, agents inferred the intended goal *g(a)* by:

$$g(a) = \underset{k}{\operatorname{argmax}} T_{(a,k)}$$

Agents then updated the model-free value of the goal, $MFG_{(g(a))}$, by:

$$MFG_{(g(a))} \leftarrow MFG_{(g(a))} + \alpha(r - MFG_{(g(a))})$$

To determine the probability of selecting action a out of choice set (a,b), agents took a weighted average W_a of the three values and entered it into a softmax function:

$$W_a = w_1 MFV_{(1,a)} + w_2 MBV_{(1,a)} + (1 - w_1 - w_2) MFG_{(g(a))}$$

$$Prob(a) = \frac{e^{\beta W_a}}{e^{\beta W_a} + e^{\beta W_b}}$$

Thus, agents were characterized by five parameters: α (the learning rate), λ (the eligibility trace), β (the softmax temperature), w_1(the model-based weight), and w_2 (the model-free weight).

References (Main text)

1 Dolan, R. J. & Dayan, P. Goals and habits in the brain. *Neuron* **80**, 312-325 (2013).

2 Thorndike, E. L. Animal intelligence: An experimental study of the associative processes in animals. *Psychological Monographs: General and Applied* **2**, i-109 (1898).

3 Cooper, R. & Shallice, T. Contention scheduling and the control of routine activities. *Cognitive Neuropsychology* **17**, 297-338 (2000).

4 Otto, A. R., Gershman, S. J., Markman, A. B. & Daw, N. D. The curse of planning: dissecting multiple reinforcement-learning systems by taxing the central executive. *Psychol Sci* **24**, 751-761, doi:10.1177/0956797612463080 (2013).

5 Huang, J. Y. & Bargh, J. A. The Selfish Goal: Autonomously operating motivational structures as the proximate cause of human judgment and behavior. *Behavioral and Brain Sciences* **37**, 121-135 (2014).

6 Lhermitte, F. 'Utilization behaviour'and its relation to lesions of the frontal lobes. *Brain* **106**, 237-255 (1983).

7 Perkins, D. N. & Salomon, G. Are cognitive skills context-bound? *Educational researcher* **18**, 16-25 (1989).

8 Adamson, R. E. Functional fixedness as related to problem solving: A repetition of three experiments. *Journal of experimental psychology* **44**, 288 (1952).

9 Sutton, R. S. & Barto, A. G. Introduction to reinforcement learning. *MIT Press*, (1998).

10 Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P. & Dolan, R. J. Model-based influences on humans' choices and striatal prediction errors. *Neuron* **69**, 1204-1215 (2011).

11 Dickinson, A., Balleine, B., Watt, A., Gonzalez, F. & Boakes, R. A. Motivational control after extended instrumental training. *Learning & behavior* **23**, 197-206 (1995).

12 McClure, S. M., Berns, G. S. & Montague, P. R. Temporal prediction errors in a passive learning task activate human striatum. *Neuron* **38**, 339-346 (2003).

13 O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H. & Dolan, R. J. Temporal difference models and reward-related learning in the human brain. *Neuron* **38**, 329-337 (2003).

14 Bayer, H. M. & Glimcher, P. W. Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron* **47**, 129-141 (2005).

15 Lashley, K. S. *The problem of serial order in behavior*. 112-131 (Wiley, 1951).

16 Botvinick, M. M., Niv, Y. & Barto, A. C. Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. *Cognition* **113**, 262-280 (2009).

17 Boyan, J. A. & Moore, A. W. Learning evaluation functions for large acyclic domains. *Proceedings of the International Conference on Machine Learning.* 63-70 (1996).

18 Zhang, N. L. & Zhang, W. Fast value iteration for goal-directed markov decision processes. *Proceedings of the Thirteenth conference on Uncertainty in artificial intelligence* 489-494 (1997).

19 Collins, A. G. & Frank, M. J. Cognitive control over learning: creating, clustering, and generalizing task-set structure. *Psychological review* **120**, 190 (2013).

20 Dayan, P. How to set the switches on this thing. *Current Opinion in Neurobiology* **22**, 1068-1074 (2012).

21 Graybiel, A. M. Habits, rituals, and the evaluative brain. *Annu. Rev. Neurosci.* **31**, 359-387 (2008).

22 O'Reilly, R. C. & Frank, M. J. Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia. *Neural Computation* **18**, 283-328 (2006).

23 Miller, E. K. The prefrontal cortex and cognitive control. *Nature Reviews* **1**, 59-65 (2000).

24 Sutton, R. S., Precup, D. & Singh, S. Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artificial intelligence* **112**, 181-211 (1999).

25 Otto, A. R., Skatova, A., Madlon-Kay, S. & Daw, N. D. Cognitive Control Predicts Use of Model-based Reinforcement Learning. *Journal of Cognitive Neuroscience* **27**, 319-333 (2014).

26 Otto, A. R., Raio, C. M., Chiang, A., Phelps, E. A. & Daw, N. D. Working-memory capacity protects model-based learning from stress. *Proceedings of the National Academy of Sciences* **110**, 20941-20946 (2013).

27 Smittenaar, P., FitzGerald, T. H., Romei, V., Wright, N. D. & Dolan, R. J. Disruption of dorsolateral prefrontal cortex decreases model-based in favor of model-free control in humans. *Neuron* **80**, 914-919 (2013).

28 Ribas-Fernandes, J. Ú. J. F. *et al.* A neural signature of hierarchical reinforcement learning. *Neuron* **71**, 370-379 (2011).

29 Daw, N. D. & Shohamy, D. The cognitive neuroscience of motivation and learning. *Social Cognition* **26**, 593-620 (2008).

References (Methods)

30 Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron* **66**, 585-595 (2010).

31 R Core Team R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/. (2014).

32 Bates, D., and Maechler, M. lme4: Linear mixed effects models using S4 classes. R package version 0.999375 33. http://CRAN.R-project.org/package=lme4. (2010).

End Notes

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Michael Frank, Samuel Gershman and Josiah Nunziato for their advice and assistance. This research was supported by grant N00014-14-1-0800 from the Office of Naval Research.

Author contributions Both authors contributed substantially to all elements of this research project.

Author information Reprints and permissions information is available at www.nature.com/reprints. The authors have no related financial interests to disclose. Correspondence and requests for materials should be addressed to cushman@fas.harvard.edu.