

Habitual goals

Adam Morris¹ & Fiery Cushman²

¹Department of Cognitive, Linguistic and Psychological Science, Brown University

²Department of Psychology, Harvard University

A rich line of decision-making research relies on the distinction between goal-directed and habitual behavior (Dolan & Dayan, 2013). When pursuing goals, people “plan ahead” over a causal model of their environment and select actions which are most likely to lead to goal fulfillment. For example, a dieter will turn down a slice of cake because he/she is pursuing the goal of weight loss. This method of action selection is flexible and powerful, but becomes computationally difficult as the causal model grows in complexity.

On the other hand, habits are simple stimulus-response patterns which get “stamped in” by reinforcement (Graybiel, 2008). For example, rats trained to press a food-releasing lever will continue to press the lever even when they are no longer hungry – pressing the lever has become a habit (Dolan & Dayan 2013). Forming habits based on reinforcement patterns is an inflexible but efficient alternative to goal-directed behavior, a way to quickly selection actions which usually lead to reward.

These systems, goal-directed and habitual, are often portrayed as competing for control of action selection (Daw et al., 2005). But we propose a cooperative interaction: goal selection itself can be habitual. There are two reasons to suspect a habitual influence on goal selection. The first is that goals exhibit habit-like properties. They can be automatically activated by contextual stimuli (Huang & Bargh, 2014), they can be unconsciously reinforced (Custers & Aarts, 2005), and they can drive behavior without conscious awareness (Huang & Bargh, 2014). These properties suggest that the system which produces stimulus-response habits might also act on goals.

The second reason is that, in a complicated world, selecting goals by searching the entire candidate goal space is computationally infeasible – there are too many candidates. For example, a person trying to satisfy their hunger would have to consider every possible goal, from turning on the TV to cleaning the bathroom to making a sandwich. Forming stimulus-goal habits based on reinforcement patterns would be a solution to this problem, a way to quickly select goals which usually lead to reward. If the goal “make a sandwich” is consistently rewarded with successful hunger reduction, it would become habitualized and automatically activate in the context of hunger.

To test for a habitual influence on goal selection, we adapted a multistep choice paradigm designed to dissociate the influence of habitual and goal-directed behavior (Glascher et al.,

2010). Participants played a decision-making game where they chose between four buttons, numbered 1-4 (Figure 1a). Buttons 1 and 3 usually led to a blue shape, and buttons 2 and 4 usually led to a red shape. But each button had a 20% chance of leading to a green shape instead. Participants received bonus points for getting different colors – blue was worth a certain amount, red a different amount, etc. Some colors won points, others lost points. But the value of each color drifted throughout the experiment, so participants had to adapt their choices to the current color values in order to maximize their winnings. On every trial, participants were presented with two out of the four buttons. They made their choice, transitioned to a colored shape, clicked on the shape, and received their bonus points.

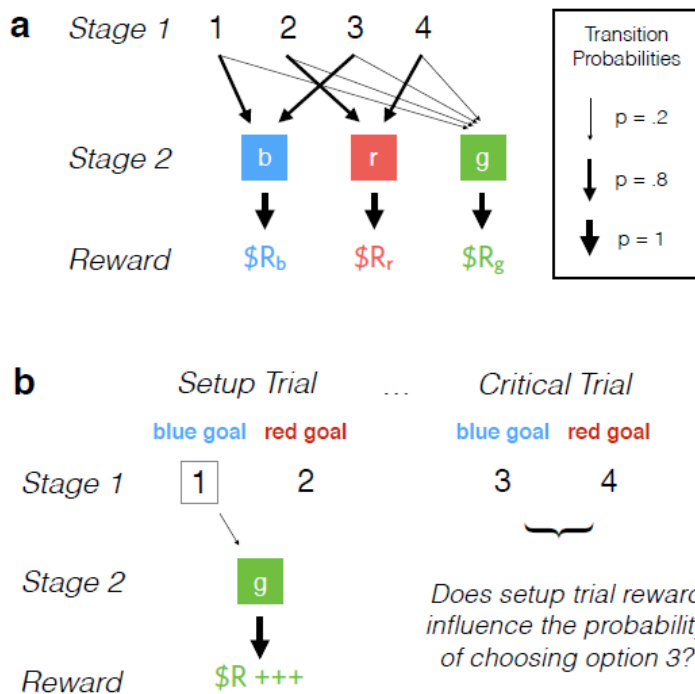


Figure 1: Task structure and logic. **a**, In Experiment 1 participants performed a two-stage decision task. In Stage 1, they were presented with two options drawn from a set of four possible options. These transitioned with variable probabilities to a set of Stage 2 shapes, which then transitioned deterministically 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 options 1 and 2 in a setup trial. Upon selecting option 1, they experience a low-probability transition to the green state and receive a large reward. A habitual influence on goal selection uniquely predicts an increase in the selection of option 3 on the subsequent critical trial, because options 1 and 3 share the common color-goal of blue.

The presence of low-probability transitions to green allowed us to isolate the influence of habitual behavior on action selection (Glascher et al., 2010). Through instruction and practice, participants were made well aware of the game's transitional structure. Specifically, they knew that each button had an equally likely chance of randomly transitioning to green – you couldn't plan to get to green. So whenever participants transitioned to green, the reinforcement they

received could not be incorporated into a forward-looking planning mechanism. On the other hand, habitual responses are blind to causal structure – they simply respond to reinforcement. So if participants become more likely to select the same action following a rewarding low-probability transition to green, and less likely following a punishing transition to green, the effect can only be due to a habitual influence on action selection.

However, we wanted to isolate a habitual influence on goal selection, not action selection. Our analysis relied on a crucial subset of trials (Figure 1b). On the “setup” trial, participants were presented with a pair of buttons – say, buttons 1 and 2. Participants selected a button, experienced a low-probability transition to green, and either received a reward or punishment. Then, on the “critical” trial, they were presented with the button that had the same color-goal as the button they just chose. For example, if the participant chose button 1, on the next trial they would be presented with button 3 (paired with either 2 or 4). 1 and 3 have the same color-goal because they both usually lead to blue.

Because the setup trial was a low-probability transition to green, a forward-planning system would not directly incorporate the reinforcement into its action selection process. Also, while a system which produces habitual actions would incorporate the reinforcement into its likelihood of selecting button 1, it would not learn anything about button 3, because choosing button 3 is a different action. So current models of the goal-directed and habitual systems predict that the reinforcement received on setup trials would have no effect on the likelihood of choosing button 3 in the critical trial.

However, if participants are forming habits of goal selection, then that reinforcement would have an effect. When selecting button 1, participants form a goal akin to “get blue”. If forming that goal is rewarded, participants would start to form a habit of selecting that goal – and because buttons 1 and 3 have the same goal, they would be more likely to select button 3 on the next trial. Similarly, if forming that goal is punished, participants would be less likely to select button 3. And that is what we find (Figure 2a). On average, participants were more likely to choose the button with the same color-goal after a rewarding trial (89% of choices) than after a punishing trial (69% of choices). The difference was significant (repeated measures t-test, $t(134) = 12.5$, $p < .0001$).

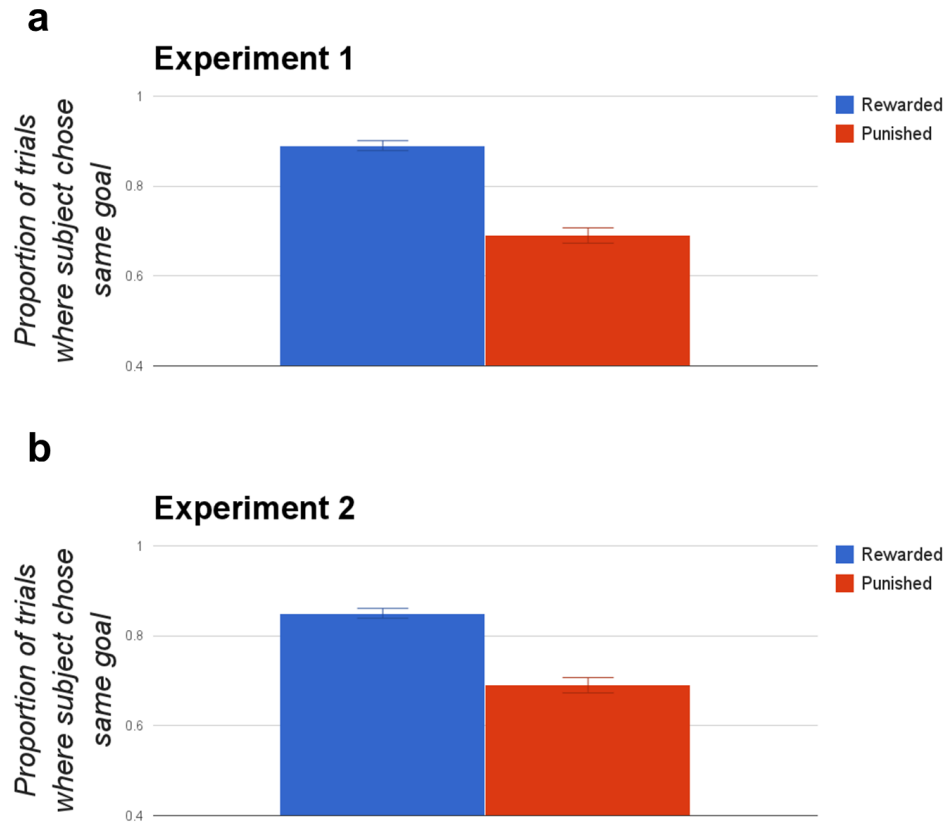


Figure 2: Results. Bars represent the proportion of critical trials on which participants chose the same color-goal, averaged across participants. Error bars indicate the standard error of the mean of these proportions across participants. **a-b** show results from Experiments 1-2, respectively. Participants were significantly more likely to choose the action with the same color-goal after reward than after punishment.

An alternative interpretation of our result is that, instead of forming bona fide habitual goals, participants were merely forming habits of “abstracted” actions – actions which somehow incorporated both buttons 1 and 3. To address this concern, we ran a second experiment in which we tested whether the representations being habitualized could be flexibly integrated into an independent causal model. This integration is a hallmark of goal-directed planning, but unlikely for “abstracted” actions.

In Experiment 2, before proceeding to the real task, participants were trained on a set of intuitive transitions from letter buttons (A,B,C,D) to number buttons (1,2,3,4). A led to 1, B to 2, etc (Figure 3a). Then the task proceeded as before, with participants choosing between number buttons, transitioning to colors, and receiving reinforcement. Except now participants were informed that, on some trials, instead of choosing between number buttons, they would choose between letter buttons (and subsequently transition to the corresponding number button). These “letter” trials always occurred on critical trials (Figure 3b). So if, on a setup trial, a participant chose button 1 and transitioned to green, on the next trial they would be presented with button C

(paired with either B or D). If participants were truly forming habitual goals, then they would be more likely to choose the letter button with the same color-goal after a rewarding critical trial than after a punishing critical trial. And that is again what we find (Figure 2b). On average, participants were more likely to choose the letter button with the same color-goal after a rewarding trial (85% of choices) than after a punishing trial (69% of choices). The difference was significant ($t(172) = -9.17, p < .0001$).

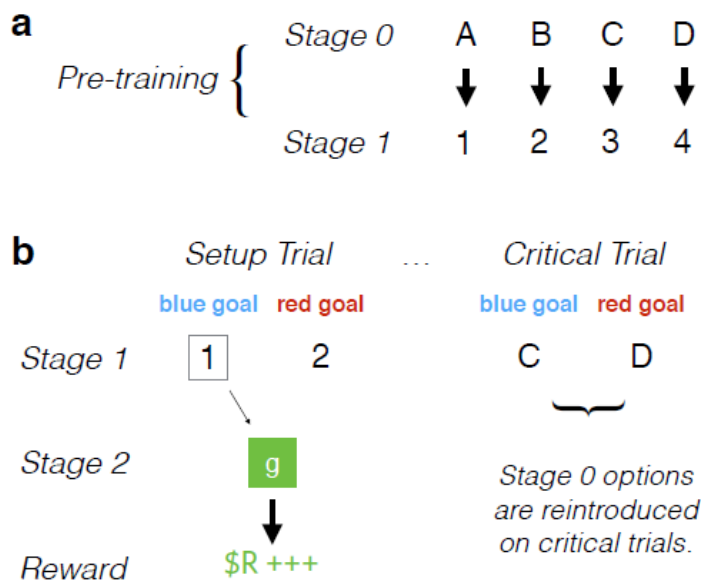


Figure 3: Experiment 2 structure and logic. Experiment 2 was modeled on the design of Experiment 1, except that **a**, participants performed a pre-training in which they learned deterministic transitions between Stage 0 letters and Stage 1 numbers, and **b**, on critical trials the Stage 0 choices were selectively reintroduced. Thus, in order to make successful choices on critical trials, participants were required to choose a Stage 0 letter that would lead to their preferred Stage 1 number.

As outlined above, our hypothesis can help explain the habit-like properties of goals, as well as the ease with which humans select goals out of a potentially infinite candidate space. Our hypothesis also has implications for a range of other issues. For example, it is debated whether complex cognitive skills are general or “context-bound” (Perkins & Salomon, 1989). The debate hinges on what it means to have acquired a cognitive skill. People often break complex cognitive tasks down into a hierarchy of goals and subgoals (Botnivick et al. 2009) – for example, long division is taught as the sequential completion of five subgoals (divide, multiply, subtract, bring down, repeat). Acquiring a cognitive skill might boil down to acquiring the proper habitual subgoal activations (“divide, then multiply, then subtract...”) in the context of higher goals (“perform long division”) or external stimuli (seeing a division problem on your test). This approach suggests a critical role for context in cognition.

In a more sinister domain, it is also debated whether drug addiction is primarily a habitual or goal-directed phenomenon (Everitt et al., 2001; Olmstead et al., 2001). Addicts’ behavior seems

to blend canonical features of habits and goals – cravings are activated automatically by contextual cues, but addicts will perform actions to obtain drugs that clearly transcend simple stimulus-response pairings. Our hypothesis offers a synthesis of these views. Addiction can be both habitual and goal-directed, with contextual cues activating the goal of obtaining drugs (which can then guide actions in a non-habitual, forward-planning way).

Finally, in the realm of moral philosophy, the Doctrine of Double Effect (DDE) is a widely invoked moral principle which states that it is wrong to harm others as a means to an end, but acceptable to harm them as a side effect (McIntyre, 2010). Many psychologists have argued that our intuitions supporting the DDE are an accidental “byproduct of cognitive architecture” (Cushman, in press), but there are competing theories for what produces those intuitions. Our hypothesis offers an elegant explanation. In order to harm somebody as a means, you have to form a goal of harming them. But forming that goal is consistently punished (Cushman, 2013). If goals can respond to reinforcement, people will end up with an aversion to forming the “harm someone” goal, making them averse to harming people as a means but not to harming them as a side effect.

References

- Botvinick, M. M., Niv, Y., & Barto, A. C. (2009). Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. *Cognition*, 113(3), 262-280.
- Cushman, F. (2013). Action, Outcome, and Value A Dual-System Framework for Morality. *Personality and social psychology review*, 17(3), 273-292.
- Cushman, F. A. (In press). The psychological origins of the doctrine of double effect. *Criminal Law and Philosophy; Special Issue on "The Means Principle."*
- Custers, R., & Aarts, H. (2005). Positive affect as implicit motivator: on the nonconscious operation of behavioral goals. *Journal of personality and social psychology*, 89(2), 129.
- Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature neuroscience*, 8(12), 1704-1711.
- Dolan, R. J., & Dayan, P. (2013). Goals and habits in the brain. *Neuron*, 80(2), 312-325.
- Everitt, B. J., Dickinson, A., & Robbins, T. W. (2001). The neuropsychological basis of addictive behaviour. *Brain Research Reviews*, 36(2), 129-138.
- Glascher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, 66(4), 585-595.

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

Perkins, D. N., & Salomon, G. (1989). Are cognitive skills context-bound?. *Educational researcher*, 18(1), 16-25.

McIntyre, A. (2010). Doctrine of double effect. *Stanford Encyclopedia of Philosophy*.

Olmstead, M. C., Lafond, M. V., Everitt, B. J., & Dickinson, A. (2001). Cocaine seeking by rats is a goal-directed action. *Behavioral neuroscience*, 115(2), 394.