Habitual goals

Adam Morris*
Department of Cognitive, Linguistic, and Psychological Science
Brown University
Providence, RI 02912
adam_morris@brown.edu

Fiery Cushman
Department of Psychology
Harvard University
Cambridge, MA 01451
cushman@fas.harvard.edu

Abstract

The distinction between habitual and goal-directed action is fundamental to decision-making research (Dolan and Dayan, 2013). Habits form as stimulus-response pairings are "stamped in" following reward. In contrast, goal-directed behavior requires planning over a causal model. Many existing models portray habitual and goal-directed systems as competing for behavioral control (Daw et al., 2005), but evidence suggests they may be codependent. Goals exhibit habit-like properties, such as automatic activation under contextual cuing (Huang and Bargh, 2014) and susceptibility to unconscious reinforcement (Custers and Aarts, 2005). Also, in complex real-world scenarios, selecting a goal out of potentially infinitely many candidates seems like an intractable problem, yet people solve it with ease – suggesting that a more efficient decision making system is influencing goal selection. We propose that goal selection can be under habitual control. Across two experiments, we demonstrate that people naturally form habitual goals which are "stamped in" by reward, but which subsequently guide behavior through model-based forward planning. The role of habitual control in goal-directed action has potential implications for a range of issues, including the contextual nature of cognitive skills, the nature of addiction, and the origin of the moral "doctrine of double effect".

Keywords: habits, goals, reinforcement learning

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.

*Corresponding author.

1 Introduction

A rich line of decision-making research relies on the distinction between goal-directed and habitual behavior (Dolan and 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 been habitualized (Dolan and 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 behavioral patterns are often formalized in the language of reinforcement learning (RL), an influential computational framework for modeling decision-making processes (Sutton and Barto, 1998). There are two broad classes of RL algorithms. Model-based algorithms maintain an internal causal model of the environment and assess actions according to their likely consequences, thus enabling goal-directed planning. In contrast, model-free algorithms do not maintain a causal model, but instead select actions according to their context-dependent history of reinforcement. The resulting stimulus-response habits are globally adaptive, but may exhibit local irrationality (Daw et al., 2011). RL models capture several core elements of human choice, including prediction-error updating in the midbrain dopamine system (Niv, 2009) and hierarchical arrangement of goals in model-based planning (Botvinick et al., 2009).

These two systems, goal-directed (model-based) and habitual (model-free), 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 habitualized. 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 and Bargh, 2014), they can be unconsciously reinforced (Custers and Aarts, 2005), and they can drive behavior without conscious awareness (Huang and 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 exhaustively searching the 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 whether people form habitualized goals, we developed a decision-making paradigm that isolates the behavioral influence of model-free goal selection. Over two experiments, we demonstrated that people spontaneously and robustly form habitual goals, which are "stamped in" by reward but guide action through model-based planning.

2 Experiments and Results

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

The presence of low-probability transitions to green allowed us to isolate a model-free influence on action selection (Gläscher 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 model-based planning mechanism. On the other hand, model-free 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 model-free influence on action selection.

However, we wanted to isolate a model-free 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

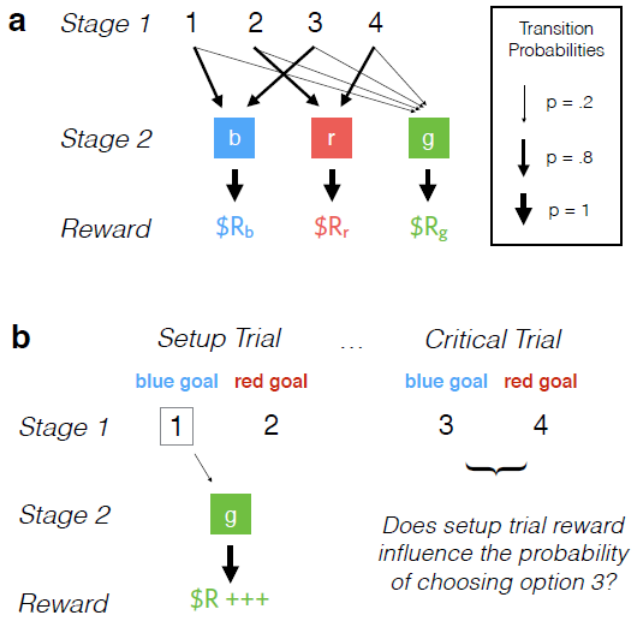

Figure 1 – Task structure and logic. a, In Experiment 1 participants performed a two-stage decision task. In Stage 1, they were presented with two options drawn from a set of four possible options. These transitioned with variable probabilities to a set of Stage 2 shapes, which then transitioned deterministically 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.

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 model-based planning system would not directly incorporate the reinforcement into its action selection process. Also, while a model-free action selection system 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, $t(134) = 12.5$, $p < .0001$).

An alternative interpretation of our result is that, instead of forming true 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

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.

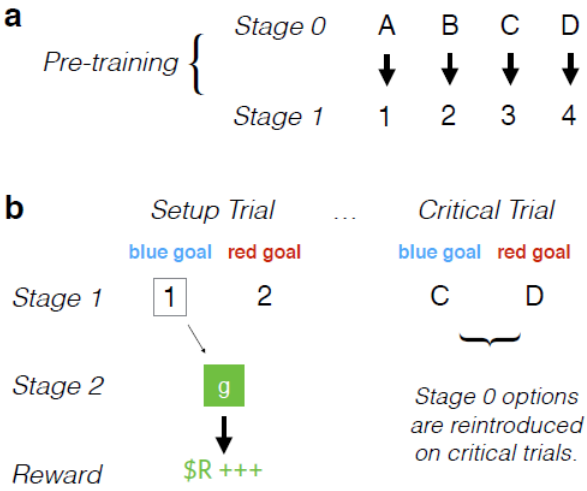

Figure 3 – Experiment 2 structure and logic. Experiment 2 was modeled on the design of Experiment 1, except that a, participants performed a pre-training in which they learned deterministic transitions between Stage 0 letters and Stage 1 numbers, and b, on critical trials the Stage 0 choices were selectively 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.

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 ($t(172) = -9.17$, $p < .0001$).

3 Discussion

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 and 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 (Botvinick 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 pernicious domain, it is also debated whether drug addiction is primarily a habitual or goal-directed phenomenon (Olmstead et al., 2001; Everitt 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 model-based 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, 2014), 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., and 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. (2014). The psychological origins of the doctrine of double effect. *Criminal Law and Philosophy*, pages 1–14.

Custers, R. and Aarts, H. (2005). Positive affect as implicit motivator: On the nonconscious operation of behavioral goals. *Journal of Personality and Social Psychology*, 89(2):129–142.

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., and Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, 69(6):1204–1215.

Daw, N. D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, 8(12):1704–1711.

Dolan, R. J. and Dayan, P. (2013). Goals and habits in the brain. *Neuron*, 80(2):312–325.

Everitt, B. J., Dickinson, A., and Robbins, T. W. (2001). The neuropsychological basis of addictive behaviour. *Brain Research Reviews*, 36(2–3):129–138.

Gläscher, J., Daw, N., Dayan, P., and 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. *Annual Review of Neuroscience*, 31(1):359–387.

Huang, J. Y. and Bargh, J. A. (2014). The selfish goal: Autonomously operating motivational structures as the proximate cause of human judgment and behavior. *Behavioral and Brain Sciences*, 37(02):121–135.

McIntyre, A. (2010). Doctrine of double effect.

Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology*, 53(3):139–154.

Olmstead, M. C., Lafond, M. V., Everitt, B. J., and Dickinson, A. (2001). Cocaine seeking by rats is a goal-directed action. *Behavioral Neuroscience*, 115(2):394–402.

Perkins, D. and Salomon, G. (1989). Are cognitive skills context-bound? *Educational Researcher*, 18(1):16–25.

Sutton, R. S. and Barto, A. G. (1998). *Introduction to Reinforcement Learning*. MIT Press, Cambridge, MA, USA, 1st edition.