Biological Psychiatry

Model-Based Control in Dimensional Psychiatry

Valerie Voon, Andrea Reiter, Miriam Sebold, and Stephanie Groman

ABSTRACT

We use parallel interacting goal-directed and habitual strategies to make our daily decisions. The arbitration between these strategies is relevant to inflexible repetitive behaviors in psychiatric disorders. Goal-directed control, also known as model-based control, is based on an affective outcome relying on a learned internal model to prospectively make decisions. In contrast, habit control, also known as model-free control, is based on an integration of previous reinforced learning autonomous of the current outcome value and is implicit and more efficient but at the cost of greater inflexibility. The concept of model-based control can be further extended into pavlovian processes. Here we describe and compare tasks that tap into these constructs and emphasize the clinical relevance and translation of these tasks in psychiatric disorders. Together, these findings highlight a role for model-based control as a transdiagnostic impairment underlying compulsive behaviors and representing a promising therapeutic target.

Keywords: Addictions, Binge eating, Compulsivity, Computational psychiatry, Goal-directed control, Habit, Model-based control, Obsessive-compulsive disorder

http://dx.doi.org/10.1016/j.biopsych.2017.04.006

We use parallel interacting goal-directed and habitual strategies to make our daily decisions, both mundane and complex. These decisions include simple ones, such as which road to drive to work, to more complex ones, such as which of multiple options to select as an investment strategy. The capacity to arbitrate between these strategies is relevant to inflexible repetitive behaviors observed in psychiatric disorders and represents a critical construct in dimensional psychiatry.

Model-based control describes a process that relies on a learned internal model of the environment to prospectively evaluate actions based on their potential outcomes (1,2). The associative structure of the model is stored and includes predictions about the consequences of each state and can be used to mentally simulate and infer values and outcomes that go beyond our previous experience. This strategy is effective and flexible, particularly with changing and novel environments, but can be computationally expensive. Goal-directed behaviors are a form of model-based control describing instrumental responding that is sensitive to the contingencies between responses and outcomes and the current value of the outcome. In contrast, model-free control describes a process in which prediction errors (what we actually receive vs. what we expected to receive) are used to estimate and store action values based on past experience. This strategy is more implicit, efficient, and rapid, with decisions based on retrospective stored values but at the cost of greater inflexibility. Habit control is a form of instrumental behavior in which responding persists despite changes in the current outcome value and represents a form of model-free control. The conceptual discrimination between flexible model-based and stored value-based behavior is commonly applied to instrumental processes using multistep tasks (1,2) but can also be applied to pavlovian processes (3). The capacity to arbitrate between these strategies is relevant dimensionally across compulsive behaviors in psychiatry.

Here we emphasize the clinical and translational relevance of goal-directed and habit control in patient populations characterized by a behavioral phenotype of compulsivity. Then we focus on the underlying neural correlates and characteristics of the two-step task highlighting commonalities and differences between humans and rodents, between task types and types of associative control.

MEASURES OF GOAL-DIRECTED AND HABIT CONTROL

Measures of goal-directed and habit control in preclinical and human studies can be divided into two basic forms: conventional overtraining and devaluation tasks (Figure 1) (4,5) and sequential decision tasks (also known as multistep tasks, with the most common being the two-step task) (Figure 2). Here we describe these concepts.

Goal-directed and habitual behaviors are common in daily decisions. We might see (stimulus) and take (response) the same turn-off when driving home (the goal). If, however, after many years of repeating this same activity (overtraining), we move to a different home, not uncommonly we will mistakenly take this old turn-off based on a habitual (overlearned stimulus–response) strategy, not taking into account the current value of the outcome (the now devalued wrong home).

Goal-directed control is governed by the knowledge of the association between actions and the value of consequences, also known as stimulus–response–outcome associations (Figure 1). With limited training of these associations, rodents

ISSN: 0006-3223
Biological Psychiatry September 15, 2017; 82:391–400 www.sobp.org/journal

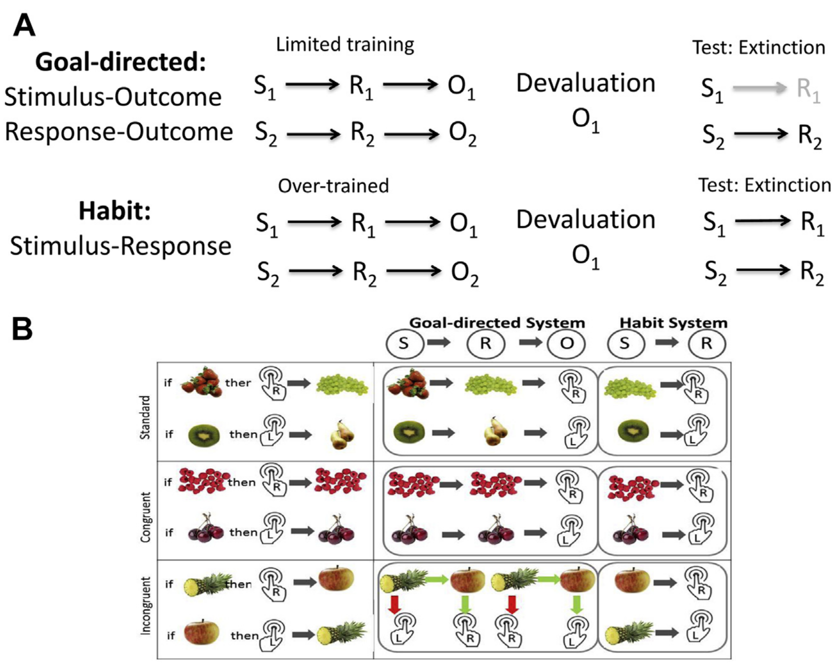

Figure 1. Conventional overtraining and devaluation tasks. **(A)** Rodents and humans undergo training to learn associations between stimulus (S), response (R), and outcome (O) contingencies. Following training, one of the outcomes is devalued (e.g., O_1). Because the subsequent test occurs under extinction (meaning that the devalued outcome is not experienced), the behavior requires access to an internal model of previous learned associations and the current value of the outcome. With moderate training, behavior is guided by stimulus–outcome or response–outcome mappings; hence, responding decreases to the devalued outcome. With extensive training, behavior becomes guided by stimulus–response mappings; hence, responding is autonomous of the current value of the outcome. For example, in this procedure, after rodents learn to obtain two different types of food, one type of food (e.g., O_1) is devalued by pairing with lithium chloride or free access to the food to induce satiety. In the probe test, a decrease in responding is normally observed to the food that is no longer valued (i.e., goal directed), but those with extensive training will persist in responding to the devalued food (i.e., habitual). **(B)** Conflict and slips-of-action task. Subjects first learn the contingencies between six cues (fruit) and responses (left [L] or right [R] button) and outcomes (fruit) for points. [Panel **B** adapted from (6). Images are from open source Stimulus Set (78).] The differentiation between goal-directed and habit learning can be assessed in one of two ways. One way is the congruent and incongruent cue–outcomes test. When the fruit cue and fruit outcome were congruent (i.e., the fruit cue leads to the same fruit outcome), both goal-directed and habitual systems were recruited, whereas only the habitual system was predominantly used when the cue and outcome were incongruent (i.e., the fruit cue leads to a different fruit outcome) because using the goal-directed system would be disadvantageous. The other way is the slips-of-action test (not shown). Instructed outcome devaluation was used to devalue two of the six fruit outcomes (i.e., subjects were told that the outcomes were no longer valuable or associated with loss of points). Subjects were then shown fruit cues in which they could earn points by pressing for the valued fruit outcome or avoid losing points by withholding pressing for the devalued fruit outcome. Habitual slips of action were characterized by pressing for the devalued fruit outcome.

remain sensitive to the current outcome value (i.e., remain goal directed). This can be assessed with devaluation of the outcome and subsequent testing of responding to learned stimuli under extinction conditions when no outcome is present. Because the test occurs without experiencing the outcome, the behavior requires access to an internal model based on previous learned associations and the current value of the outcome. With overtraining of these associations, sensitivity to the current value of the outcome is decreased with increased reliance on stimulus–response associations (i.e., shifts toward habit). Thus, responses become persistent and fail to shift flexibly with changes in current outcome value (4,5).

Human studies have similarly translated overtrained and devaluation tasks. One specific design uses overtraining and testing with a conflict procedure and "slips of action" to assess goal-directed and habit control (6) (Figure 1).

Multistep tasks based on reinforcement learning models have been applied to goal-directed and habit control, also known as model-based and model-free control (1,2,7). The two-step task is a sequential two-stage decision task in which subjects choose between one of two choices at each state, leading in the second stage to a rewarded or nonrewarded outcome of varying probability (1) (Figure 2). Choices at the first stage are associated with a likely transition and an unlikely transition of fixed probability to one of two states. Model-free habitual control is based on the repetition of a previously rewarded action irrespective of the likelihood of the transition, whereas model-based goal-directed control takes into account the task model and the likelihood of the transition. The task provides an index of the relative balance between model-based and model-free control (Figure 2).

NEURAL SUBSTRATES OF GOAL-DIRECTED AND HABIT TASKS IN HEALTHY HUMANS

Rodent and human studies implicate similar dissociable frontostriatal regions in the balance between goal-directed and habit learning. Lesions of the rodent dorsomedial striatum (human caudate) and prelimbic cortex block goal-directed behaviors, leaving intact habit learning (8,9). In contrast, lesions of the dorsolateral striatum (human putamen) and infralimbic cortex result in intact goal-directed behaviors despite extended training (9–11).

Human functional magnetic resonance imaging studies using overtrained and devaluation tasks show a clear dissociation: goal-directed behaviors are associated with ventromedial prefrontal cortex (vmPFC) and caudate activity, regions implicated in action–outcome encoding and outcome valuation relevant to tracking immediate outcome values, and habit learning is associated with putaminal regions. Following training on reinforcement learning tasks, greater habitual behaviors over the course of learning were associated with increased posterior putaminal activity (12) and greater