Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Behavioral Sciences

What are habits and do they depend on the striatum? A view from the study of neuropsychological populations

Karin Foerde

What are the neural substrates of habit learning in humans? Studies in neuropsychological populations have been central to answering the question, and for decades, research appeared to have provided a fairly consistent answer. However, developments in assays of habits in animals, as well as new approaches to dissecting habitual versus goal-directed control of behavior in humans, point to further complexities in human habit learning. This has raised new questions about the status of habits in neuropsychological populations and our understanding of how the brain supports habitual behavior. I review these emerging challenges and suggest a more nuanced approach to habit learning.

Address
Department of Psychology, University of Amsterdam, Netherlands

Corresponding author: Foerde, Karin (K.E.G.Forde@uva.nl)

Current Opinion in Behavioral Sciences 2018, **20**:17–24

This review comes from a themed issue on **Habits and skills**

Edited by **Barbara Knowlton** and **Jörn Diedrichsen**

http://dx.doi.org/10.1016/j.cobeha.2017.08.011

Introduction

Habits are critical for adaptive behavior: They allow organisms to capitalize on invariance in the environment while directing valuable cognitive resources toward more demanding or exciting endeavors [1]. Yet, an over-reliance on habits renders behavior inflexible in the face of change, leading instead to inappropriate behaviors. Indeed, the concept of *habit* has been central to how we understand of a broad range of behaviors that are both adaptive and maladaptive (e.g., addiction [2], OCD [3], and eating disorders [4]).

But how coherent is the habit concept? Based on studies in neuropsychological patients, the neural bases of habit learning in humans were thought to be well established. Yet, as data from these populations accumulate, it is becoming clear that a single construct cannot explain the pattern of observed impairments. These findings may call into question whether habits are truly impaired in some neuropsychological populations. Alternatively, they suggest it may be necessary to distinguish between multiple different forms of habits. I consider these possibilities in light of emerging evidence from neuropsychology, highlight key challenges, and suggest some steps toward an updated integrative framework.

What are we studying when we study habits?

The habit construct is often defined by a collection of attributes: (a) habits are learned gradually rather than being innate, (b) learning and performance can proceed without full attention (i.e., under distraction), (c) learned behavior can be performed automatically, potentially without conscious awareness of what was learned, (d) sequences of simple behaviors may become routinized and performed as single units of behavior, and (e) habitual behavior is inflexible and becomes insensitive to the outcomes of behavior after extensive training [1,5,6]. Generally, stimulus–response (S–R) associations are thought to underlie habit formation [7,8]. However, the habit characteristics do not always cluster together; some may emerge without others [9]. For example, when habit acquisition is followed by outcome devaluation, continued responding for the devalued outcome (the mark of habitual behavior) may or may not be accompanied by awareness of correct response-outcome associations or the change in outcome value. This divergence between attributes results in a tension between efforts to precisely operationalize habits for study and to sufficiently capture the broad phenomenon.

Habit learning in neuropsychological populations — the human lesion model

Much of what we know about habit learning in humans comes from studies of patients with brain damage. Seminal discoveries revealed that damage to the hippocampus and medial temporal lobes (MTL) caused dense amnesia but spared a variety of learning capacities [7,10], some of which shared the features of habits described above. In addition, based on animal lesion studies, it was suggested that behavior is dependent on habit formation following hippocampal damage [7]. Thus, habits were initially defined in terms of capacities that functioned independently of the hippocampus.

The striatum was first described as the primary substrate of habits by Mishkin *et al.* [11] — a proposal that has been supported by a wealth of lesion studies in animals (rats and monkeys) [8,12–16]. Thus, to study habit learning in humans, researchers turned to patient populations with

striatal damage. The most commonly studied patients are individuals with Parkinson's disease (PD) or Huntington's disease (HD). PD patients have striatal dysfunction due to loss of dopaminergic input from the midbrain to the striatum [17–19], with detrimental effects on learning, memory, and cognition, as well as motivation and affect [20–22]. Dopamine in the striatum conveys critical learning signals [23,24] and underlies learning-dependent plasticity [25]. In HD, striatal dysfunction is the direct result of atrophy and neuronal loss in the striatum [26] rather than loss of dopamine. PD and HD patients are typically studied during early disease stages when dysfunction is relatively restricted to dorsal striatal regions (the putamen in PD; putamen and caudate in HD) as opposed to ventral regions (nucleus accumbens) [17–19,26–28], providing a reasonable analog to lesion studies in animals (Figure 1).

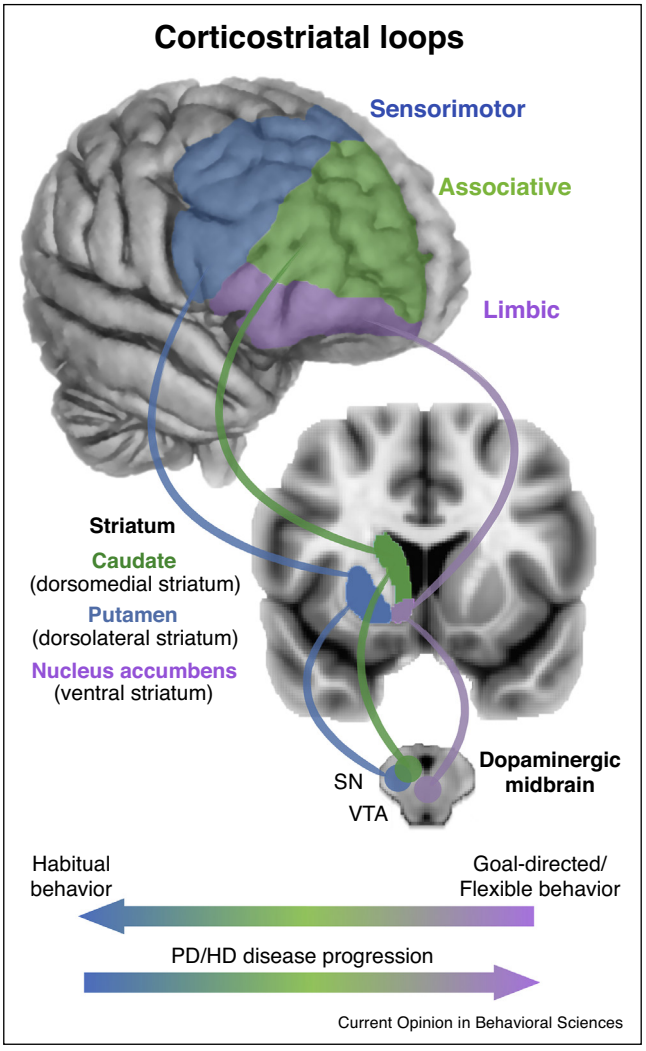
Investigating habit learning in humans

Having identified an appropriate human lesion model, the next significant challenge was to devise appropriate tests of habit learning. A simple solution might be to directly adapt paradigms used in animals, but the risk is that humans and non-human animals may not learn tasks using the same cognitive and neural mechanisms [16,29,30]. Consider the *concurrent discrimination task*: animals learn the correct responses gradually from feedback over hundreds of trials, but healthy adults learn the same responses in just a few trials, and have flexible, declarative knowledge about their learning. Interestingly, Bayley *et al.* [31•] tested two amnesic patients with no declarative memory and found that they learned the task, but did so in a gradual, rote, and inflexible manner. This suggested that while humans can learn S–R habits gradually, like rodents and monkeys, they do so only when declarative memory abilities are inactivated. Such findings reveal the need for tasks that capture gradual S–R habit learning while circumventing rapid, declarative (and MTL dependent) memory.

Perhaps the most commonly-used task thought to achieve this goal is the probabilistic classification task (PCT) [32], which is thought to capture the incremental learning from response-contingent feedback characteristic of tasks used in animals. The PCT requires participants to make binary predictions based on a complex set of cues that are probabilistically associated with outcomes (see Figure 2a, b and caption for detailed task description). The complexity of the task structure is intended to overwhelm the ability to rapidly memorize cue-outcome associations. Indeed, Collins *et al.* recently showed that when working memory was challenged by an increased number of stimuli, participants relied more on gradual S–R learning mechanisms [33•,34]. As expected, studies showed that whereas MTL damage did not prevent learning of the PCT [32,35,36], PD and HD patients were impaired at learning despite having intact declarative task knowledge [35,37–39]. Thus, studies using the PCT in neuropsychological populations confirmed the prediction from animal studies that human habit learning depends on the striatum and not the MTL.

Another task used extensively to investigate habits and skills is the serial reaction time task (SRTT; [40]), which involves making rapid motor responses to cues that, unbeknownst to the participant, follow a repeating sequence (see Figure 2c,d and caption for detailed task description). The SRTT captures a different aspect of habits than the PCT — the SRTT assesses the development of chunked representations of series of behaviors

Figure 1



Simplified schematic of three corticostriatal loops important for habitual and goal-directed behavior. (Bottom) Hypothesized progression of involvement of corticostriatal loops as behavior develops from goal-directed to habitual and progression of striatal dysfunction in neuropsychological populations. VTA, ventral tegmental area; SN, substantia nigra; PD, Parkinson's disease; HD, Huntington disease.