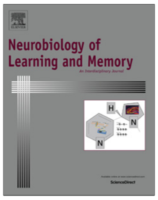
Neurobiology of Learning and Memory xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Neurobiology of Learning and Memory

journal homepage: www.elsevier.com/locate/ynlme



Invited Review

A universal role of the ventral striatum in reward-based learning: Evidence from human studies

Q1 Reka Daniel [a,b,*], Stefan Pollmann [a,c]

[a] *Department of Experimental Psychology, Otto-von-Guericke-Universität Magdeburg, D-39016 Magdeburg, Germany*
Q2 [b] *Princeton Neuroscience Institute, Princeton University, Green Hall, Princeton, NJ 08540, USA*
[c] *Center for Behavioral Brain Sciences, D-39016 Magdeburg, Germany*

ARTICLE INFO

Article history:
Received 6 December 2013
Revised 1 May 2014
Accepted 3 May 2014
Available online xxxx

Keywords:
Ventral striatum
Reward
Feedback
Learning
fMRI
Human

ABSTRACT

Reinforcement learning enables organisms to adjust their behavior in order to maximize rewards. Electrophysiological recordings of dopaminergic midbrain neurons have shown that they code the difference between actual and predicted rewards, i.e., the reward prediction error, in many species. This error signal is conveyed to both the striatum and cortical areas and is thought to play a central role in learning to optimize behavior. However, in human daily life rewards are diverse and often only indirect feedback is available. Here we explore the range of rewards that are processed by the dopaminergic system in human participants, and examine whether it is also involved in learning in the absence of explicit rewards. While results from electrophysiological recordings in humans are sparse, evidence linking dopaminergic activity to the metabolic signal recorded from the midbrain and striatum with functional magnetic resonance imaging (fMRI) is available. Results from fMRI studies suggest that the human ventral striatum (VS) receives valuation information for a diverse set of rewarding stimuli. These range from simple primary reinforcers such as juice rewards over abstract social rewards to internally generated signals on perceived correctness, suggesting that the VS is involved in learning from trial-and-error irrespective of the specific nature of provided rewards. In addition, we summarize evidence that the VS can also be implicated when learning from observing others, and in tasks that go beyond simple stimulus-action-outcome learning, indicating that the reward system is also recruited in more complex learning tasks.

1. Introduction

As any living organism, humans are faced with the need to make decisions about how to act in response to a plethora of environmental cues every day. Often, we encounter similar situations repeatedly, which enables us to use past experiences to predict future outcomes (Cohen, 2008). Learning from trial and error, or reinforcement learning, has been extensively studied in the last decades. Normative computational models have proven to be successful in explaining learning in terms of a reward prediction error, i.e., a mismatch between predicted and actual rewards. On a neural level, it has been shown that the dopaminergic midbrain neurons, with their massive projections to the ventral striatum (VS), represent this reward prediction error and play a central role in reward-based learning (for reviews see Delgado (2007), Niv and Montague (2008), O'Doherty, Hampton and Kim (2007), and Schultz (2007)).

However, in ecologically valid settings, decision problems vary hugely in complexity: organisms may have to decide which objects in the environment to categorize as nutrition, but also which partner to choose, or, in the human case, what career path to follow, and how to provide for retirement. These problems also vary in the rewards they provide: from simple primary and secondary reinforcers such as food, the option to reproduce, and money, to more abstract rewards such as love, social approval and (financial) stability. For many of these decisions, feedback from the environment is sparse and delayed (Hogarth, 2006). In addition, learning from errors is costly, and their commission is often avoidable. In these situations we have to additionally rely on other mechanisms than learning from trial and error, such as building abstract structural representations (Diuk, Tsai, Wallis, Botvinick, & Niv, 2013; Ribas-Fernandes et al., 2011), mental models of the environment (Daw, Gershman, Seymour, Dayan, & Dolan, 2011; Gl_scher, Daw, Dayan, & O'Doherty, 2010) or learning from others (Burke, Tobler, Baddeley, & Schultz, 2010). Many species are able to learn from such indirect experience, i.e., by observing the outcome of others and imitating their actions (Chamley, 2004). Humans in particular

---

* Corresponding author at: Princeton Neuroscience Institute, Princeton University, Green Hall, Princeton, NJ 08450, USA. Tel.: +1 609 258 5456.
E-mail address: rekad@princeton.edu (R. Daniel).

http://dx.doi.org/10.1016/j.nlm.2014.05.002
1074-7427/© 2014 Published by Elsevier Inc.

Please cite this article in press as: Daniel, R., & Pollmann, S. A universal role of the ventral striatum in reward-based learning: Evidence from human studies. *Neurobiology of Learning and Memory* (2014), http://dx.doi.org/10.1016/j.nlm.2014.05.002

have developed a complex form of communication, allowing them to pass on learned information by providing instructions (Li, Delgado, & Phelps, 2011).

The neural substrates of learning from abstract, incomplete or absent rewards are only currently being investigated. Humans are the ideal species to study these tasks: they are willing to work without explicit rewards after each trial, which facilitates the study of observational learning, and can be verbally instructed to examine the influence of prior knowledge. In addition, experimental methodology to study higher cognitive functions is well established, and a wealth of experimental paradigms, formal models and empirical behavioral data exists on learning in humans. Using functional magnetic resonance imaging (fMRI) it is possible to non-invasively study the brain activation of human participants performing complex learning and decision making tasks.

This review focuses on how central findings on the neural underpinnings of animal learning can be replicated and elaborated upon by examining different types of reward and more complex forms of learning in human participants. To this end, we present a summary of influential results in the animal literature, and discuss how activation data obtained using fMRI relates to more direct measures of neural firing as acquired in electrophysiological recordings, and specifically to dopaminergic activity. We then go on to summarize recent findings on the neural substrates of learning in humans, and argue that phylogenetically old pathways that mediate simple stimulus-response learning are recruited even in the absence of explicit rewards to solve complex decision making tasks.

2. Reinforcement learning and the dopaminergic system

In order to optimally behave in any given environment it is crucial to determine which actions result in rewarding events given a specific state of the environment. These rewarding events come in many different shapes and flavors: even primary rewards can be as diverse as the opportunity to mate or a drop of juice. Any agent striving to maximize positive outcomes can benefit from a system that unifies these diverse sensory inputs by encoding the motivational properties of stimuli, thereby providing a “retina of the reward system” (Schultz, 2007, 2008).

2.1. The dopaminergic system: A system for reward and motivation

The seminal experiments of Olds and Milner (1954) were a first step towards identifying a system in the mammalian brain that is dedicated to processing motivational information. They described several brain sites in the rat where direct electrical stimulation acted as a reinforcer, inciting the animal to stimulate itself. Subsequent experiments showed that the tissue inducing the highest rates of self-stimulation is located in the medial forebrain bundle (MFB), which is connected to dopamine cell bodies in the ventral tegmental area (VTA) (Gallistel, Shizgal, & Yeomans, 1981), and that self-stimulation of the MFB is related to fluctuations in dopamine levels in one of the prominent targets of mesolimbic dopamine neurons, the nucleus accumbens (Garris, Kilpatrick, & Bunin, 1999; Hernandez et al., 2006). Gallistel et al. (1981) point out three central properties of behavior during self-stimulation: rats engage in self-stimulation with high vigor, i.e., a short latency and high intensity of responding. In addition, they pursue self-stimulation over other vital goals, and adjust their behavior flexibly to the magnitude of recent stimulation. These three properties can be viewed as hallmarks of motivation: motivation energizes the organism, directs its behavior towards a goal, and enables learning about outcomes (Gallistel et al., 1981).

2.2. The dopaminergic system in learning

Dopaminergic neurotransmission has been associated with a wide variety of functions, many of which are important for optimizing behavior, i.e., for maximizing reward and minimizing aversive outcomes. Different firing modes of dopaminergic neurons have been associated with the facilitation of a wide range of motor, cognitive, and motivational processes (Bromberg-Martin, Matsumoto, & Hikosaka, 2010; Schultz, 2008), with determining the strength and rate of responding (Niv, Daw, Joel, & Dayan, 2007), and with signaling different aspects of salience, including appetitive and aversive information (Bromberg-Martin et al., 2010; Schultz, 2007), novelty (Bunzeck & D_zel, 2006; Ljungberg, 1992; Wittmann, Bunzeck, Dolan, & D_zel, 2007), and contextual deviance (Zaehle et al., 2013). However, the majority of recent research has focused on one specific aspect of the firing of dopaminergic neurons: 75–80% of the cells convey a signal that is ideally suited to promote learning (Schultz, 2008). They respond with short-latency phasic bursts to unpredicted rewards and reward-predicting stimuli. Importantly, when rewards are fully predicted no response is observed, while the omission of predicted rewards leads to a decrease in baseline activity (Schultz, 2002) (see Fig. 1A). This behavior corresponds to a reward prediction error signal, i.e., to a coding of reward information as the difference between received and expected reward (Montague, Dayan, & Sejnowski, 1996; Schultz, Dayan, & Montague, 1997).

A reward prediction error signal has also been postulated as a teaching signal in normative theories of reinforcement learning (Bertsekas & Tsitsiklis, 1996; Sutton & Barto, 1990). These theories focus on providing normative accounts of how agents can optimize their behavior (Niv & Montague, 2008). A wide variety of reinforcement learning algorithms exists, however most of them share some core features: they predict that whenever faced with a decision, the agent calculates a value for each available option. In order to allow for random and exploratory behavior, these values are subsequently passed through a probabilistic function before the stimulus or action with the highest value is chosen. Whenever new information becomes available, e.g. in the form of an unpredicted reward or the omission of an expected reward, the values are updated using the reward prediction error multiplied by a learning rate (Cohen, 2008).

Such a mechanism for calculating the expected subjective value of future states of the environment and updating it using a simple prediction error signal would benefit any living organism: it provides a mechanism to make predictions based on the similarity between past experiences and the current state of the environment, enabling the organism to make informed decisions. Whenever these predictions are violated, they can be updated using a single scalar signal, without having to (re-)process all available sensory information (Schultz, 2008). A correspondence between the prediction error signals postulated by normative models and phasic dopaminergic firing was shown quantitatively (Bayer & Glimcher, 2005; Tobler, Fiorillo, & Schultz, 2005) and using several different paradigms, including blocking (Waelti, Dickinson, & Schultz, 2001) and conditioned inhibition (Tobler, Dickinson, & Schultz, 2003). This postulated role of short-latency firing of dopaminergic midbrain neurons in signaling prediction errors has been termed the reward prediction error hypothesis of dopamine (for extensive reviews see Schultz (2002, 2006)).

One potential mechanism by which phasic dopamine bursts could support learning is suggested by the observation that long-term potentiation in the striatum depends not only on strong pre- and postsynaptic activation, but also on dopamine release (this is sometimes referred to as the three-factor learning rule (Ashby & Ennis, 2006)). According to this model, the same dentritic spines of striatal medium spiny cells are contacted both by

Please cite this article in press as: Daniel, R., & Pollmann, S. A universal role of the ventral striatum in reward-based learning: Evidence from human studies. *Neurobiology of Learning and Memory* (2014), http://dx.doi.org/10.1016/j.nlm.2014.05.002