Neurobiology of Learning and Memory 155 (2018) 486–497

Contents lists available at ScienceDirect

Neurobiology of Learning and Memory

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

Review

Control of habitual instrumental actions by anticipation of postingestive sensations

Denis Drapeau

Pelham Process Sciences Institute, 150 Bridge Street #779, Pelham, NH 03076, USA

ARTICLE INFO

Keywords:
Motivational state
Chunking
Reinforcement learning
Irrelevant incentive effect
Sensory-specific satiety
Dopamine

ABSTRACT

In animal behavioral experiments, extended training causes instrumental actions that deliver ingestible substances to lose sensitivity to outcome devaluation and contingency degradation, and to gain direct sensitivity to the current motivational state. These features of habitual control have been attributed to a process that relies on stimulus-response (S-R) associations linking the context to instrumental actions and Pavlovian associations linking the same context to orally-sensed properties of substances attained there. This Pavlovian process was conceived based on the results of irrelevant incentive experiments, but it is not supported by the results of all such experiments. An alternative process is therefore proposed here. In this process, recall of the instrumental action is evoked by an S-R association, but the rate at which this action is then performed is controlled by anticipation of postingestive sensations that have frequently followed it. This anticipation relies on recall of an association linking the action directly to the postingestive sensations. This association is learned during the formation of a chunked action series that begins with the instrumental action and ends with a consummatory response. It enables a prediction of subjective value that is directly influenced by the current motivational state, but is not influenced by devaluation or non-contingent delivery of the substance that has produced the postingestive sensations.

1. Introduction

In animal behavioral experiments, control of an instrumental action that delivers an ingestible substance tends to use the goal-directed process early in training, but shifts to a habitual process during extended training. The goal-directed process displays sensitivity to both outcome devaluation and contingency degradation (Dickinson & Balleine, 1994), whereas the habitual process lacks these sensitivities. Interventions that interfere with the functioning of the goal-directed process likewise cause control to lack sensitivity to outcome devaluation and contingency degradation. Such interventions include bilateral inactivation of the gustatory cortex (Balleine & Dickinson, 2000), dorsomedial striatum (DMS) (Yin, Ostlund, Knowlton, & Balleine, 2005; Corbit & Janak, 2010), the basolateral amygdala (BLA) (Corbit & Balleine, 2005; Shiflett & Balleine, 2010), the mediodorsal thalamus (Corbit, Muir, & Balleine, 2003), or the prelimbic cortex (Corbit & Balleine, 2003; Tran-Tu-Yen, Marchand, Pape, Di Scala, & Coutureau, 2009; Hart, Bradfield, & Balleine, 2018).

1.1. The goal-directed process

The goal-directed process relies on two distinct associations, each of which is learned during instrumental training (Dickinson & Balleine, 1994). One association is an “action-outcome contingency”. This association represents the likelihood a specific instrumental action will deliver a specific outcome, where “outcome” refers to a set of external conditions represented by exteroceptive visual, auditory, touch, odor, and taste sensations. The other association is an “outcome value” that links the specific outcome to a subjective value. When a specific outcome is recalled by the animal, these two associations together generate a prediction of the subjective value of the action. This prediction then controls the action’s likelihood or rate.

Sensitivity to outcome devaluation, as well as loss of such sensitivity, is illustrated by an experiment described by Holland (2004). In an operant conditioning chamber, food-restricted rats learned to press a lever that delivered food pellets. Then the pellets were devalued for the “paired” group. This was achieved by offering free access to pellets in the home cages on each of two days, with each free access period being followed by an injection of lithium chloride. Lithium chloride injections following ingestion of a food can produce a subsequent reluctance to seek or ingest that food (Adams & Dickinson, 1981). On the test day, the rats were returned to the chamber, and their baseline rates of lever pressing were evaluated in extinction (prior to the introduction of Pavlovian cues for the assessment of Pavlovian-instrumental transfer).

E-mail address: drapeau@alumni.hmc.edu.

https://doi.org/10.1016/j.nlm.2018.09.011
Received 21 May 2018; Received in revised form 14 August 2018; Accepted 26 September 2018
Available online 27 September 2018
1074-7427/

If the instrumental training had included only 125 action-outcome pairings, the rats showed sensitivity to outcome devaluation, pressing at a rate only half as high as rats in an "unpaired" group for whom the free access periods had not been followed by lithium chloride injections. Rats in both groups had learned an action-outcome contingency linking lever pressing to delivery of food pellets, and had also learned an outcome value linking food pellets to favorable subjective value. For rats in the paired group, the outcome value association had then been updated so that it also represented the unfavorable subjective value produced by the lithium chloride. As a result, when these rats were again offered the lever, they were less inclined to press it.

In contrast, if the instrumental training had included 500 action-outcome pairings, rats in the paired group pressed at the same rate as rats in the unpaired group (Holland, 2004, Fig. 2A). Thus the additional action-outcome pairings had caused control of lever pressing to lose sensitivity to devaluation of the pellets.

Sensitivity to contingency degradation, as well as loss of such sensitivity, is illustrated by an experiment described by Balleine, Killcross, and Dickinson (2003). Food-restricted rats were trained to perform two instrumental actions (lever press and chain pull), one that delivered food pellets and one that delivered maltodextrin solution. During each 1-sec interval, the first performance of either action delivered the corresponding food with a 5% likelihood. Then one of the two foods became no longer contingent on the corresponding action, being delivered with this same 5% likelihood during each 1-sec interval regardless of whether the action had been performed. Rats in the control group, which were using the goal-directed process, as indicated by sensitivity to outcome devaluation, progressively decreased their rate of performing the action that had delivered the now non-contingent food. In contrast, rats with bilateral BLA lesions did not display sensitivity to outcome devaluation, and also did not decrease their rate of performing the action that had delivered the now non-contingent food.

1.2. The habitual process

The habitual process is generally assumed to be a stimulus-response (S-R) process. For example, Balleine and O'Doherty (2010, p. 49) stated that "actions under goal-directed control are performed with regard to their consequences, whereas those under habitual control are more reflexive in nature, by virtue of their control by antecedent stimuli rather than their consequences". In an S-R process, the inclination of an agent to perform a specific action is based on the strength of a link connecting the stimulus to the action. This strength develops when the stimulus is frequently followed by the action due to favorable value of a subsequent consequence.

Control by an S-R process would explain the lack of sensitivity to outcome devaluation displayed by rats in the Holland (2004) experiment following 500 action-outcome pairings, because neither the stimulus (the chamber context) nor the action (pressing the lever) was present during the devaluation procedure. The strength of the S-R link would not get updated until the devalued outcome had occurred following the stimulus and the action. Control by an S-R process would also explain the lack of sensitivity to contingency degradation by BLA-lesioned rats in the Balleine et al. (2003) experiment, because the strength of the S-R link would not be affected by additional non-contingent occurrences of the outcome.

However, a shift to an S-R process is not the only behavioral change that tends to occur during habit formation (Graybiel, 1998, 2008; Seger & Spiering, 2011; Smith & Graybiel, 2016; Robbins & Costa, 2017). Another change is "chunking", which is the merging of the instrumental action with one or more subsequent actions to form a chunked series (Graybiel, 1998; Jin, Tecuapetla, & Costa, 2014; Smith & Graybiel, 2014, 2016). After chunking has occurred, the instrumental action is preceded by a burst of striatal activity, and subsequent actions in the chunked series are preceded by much less striatal activity (Jog, Kubota, Connolly, Hillegaart, & Graybiel, 1999; Barnes, Kubota, Hu, Jin, & Graybiel, 2005). This suggests that the prediction of value that precedes the instrumental action is based on anticipation of the outcome of the whole chunked series, rather than anticipation of the immediate outcome. Chunking is therefore an alternative explanation for the loss of sensitivity to outcome devaluation and contingency degradation (Graybiel, 1998; Dezfouli, Lingawi, & Balleine, 2014). The action can still be goal-directed, but the goal will be the outcome of the chunked series rather than the immediate outcome.

1.3. Direct sensitivity to the motivational state

Although the goal-directed process enables control of an instrumental action to be sensitive to both outcome devaluation and contingency degradation, it does not enable it to be directly sensitive to the motivational state. The motivational state is the set of internally sensed extents of deprivation with respect to conditions that are important for survival or reproduction. Specific "dimensions" of the motivational state include the extents of water and sodium deprivation, as well as various representations of the extent of food deprivation. Direct sensitivity to the motivational state is sensitivity to a new motivational state even when the outcome of the action has never been experienced under that motivational state. Although the goal-directed process does not enable direct sensitivity to the motivational state, it enables "incentive learning", which is learning of the effect of a new motivational state on the outcome value (Dickinson & Balleine, 1994).

Incentive learning is illustrated by an experiment by Dickinson, Balleine, Watt, Gonzales, and Boakes (1995). The "120 group" of rats in this experiment had learned to press a lever that delivered food pellets during 120 action-outcome pairings while food-restricted. When tested in extinction, rats in this group pressed the lever at the same rate when sated as when food-restricted unless they had previously experienced the pellets while sated, in which case they pressed it at a lower rate. During this previous experience, they had apparently learned that the outcome value of pellets is lower under the sated motivational state.

Interestingly, an overtrained group of rats in the Dickinson et al. (1995) experiment that had experienced 360 action-outcome pairings displayed *direct* sensitivity to the motivational state. Rats in this group pressed the lever at a lower rate when sated even if they had not previously been exposed to the pellets while sated (Dickinson et al., 1995, Fig. 1; Balleine & Dickinson, 1998, p. 412). Additional examples reviewed below in Section 2 indicate that extended training, as well as interventions that interfere with the functioning of the goal-directed process, consistently cause control of instrumental actions that deliver foods to become directly sensitive to food satiety. Control thus seems to shift from the goal-directed process, which displays sensitivity to both outcome devaluation and contingency degradation but lacks direct sensitivity to the motivational state, to a habitual process that lacks sensitivity to outcome devaluation and contingency degradation but displays direct sensitivity to the motivational state.

The obvious explanation for this gain of direct sensitivity would be the learning of an association linking the instrumental action to orally-sensed properties of the substance it delivers. A link to sweetness would enable the action to be directly sensitive to food-deprivation, a link to wetness would enable the action to be directly sensitive to water deprivation, and a link to saltiness would enable the action to be directly sensitive to sodium deprivation. However, this explanation is not supported by results of the "irrelevant incentive" experiments reviewed below in Section 3.1. In an irrelevant incentive experiment, instrumental training occurs under one motivational state and testing occurs under a new motivational state. In the experiments reviewed in Section 3.1, the rate of a specific instrumental action was affected by a new motivational state when this motivational state should not have affected the subjective value of the substance delivered by the action, but should have affected the subjective value of another substance that had been frequently attained in the same context.

These results have been attributed to a Pavlovian process (Dickinson