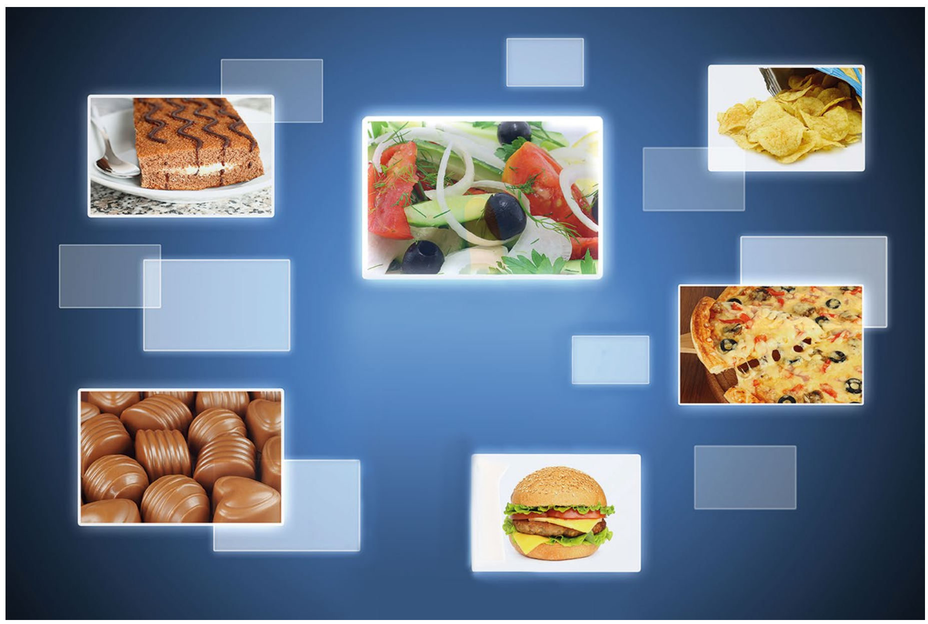
NEUROSCIENCE

More than two forms of Pavlovian prediction

Behavioural neuroscience and reinforcement learning theory distinguish between 'model-free' and 'model-based' computations that can guide behaviour. A recent study demonstrates that Pavlovian learning can give rise to behavioural responses that are not well accounted for by this existing dichotomy, suggesting that there may be greater complexity to the computations that underlie Pavlovian prediction.

Hillary A. Raab and Catherine A. Hartley

Pavlovian learning enables cues that predict positive and negative outcomes to elicit reflexive behavioural responses. Traditionally, this type of learning has been proposed to stem from computations that transfer positive or negative value from a valenced outcome to a predictive cue[1]. However, recent empirical findings call this account into question, demonstrating that Pavlovian conditioned responses not only reflect the motivational value of a predicted reward or punishment but can also reveal specific knowledge about that outcome (for example, when or where it will happen, or how it will look, taste, or feel)[2,3]. Such evidence that conditioned responses can recruit a mental model of specific outcome expectations has motivated the suggestion that Pavlovian learning not only arises from a model-free value-transfer process but can also engage underlying model-based computations[4]. Challenging this dual-systems framework, a study by Pool et al. published in *Nature Human Behaviour* presents evidence of learned Pavlovian responses that do not conform to the defining properties of either model-free or model-based evaluation[5].



Credit: Rvlsoft / Alamy Stock Photo

The distinction between model-free and model-based evaluation has been an area of extensive study in the domain of instrumental learning (i.e., determining which actions are good or bad)[6]. Model-free algorithms use prior experience with rewards or punishments to update a stored estimate of the average value of action ("going to Coffee Place A is good") but maintain no other representation of predicted outcomes and their specific features. In contrast, model-based algorithms use representations of environmental contingencies ("to get to Coffee Place A, I walk to the next corner") and the specific features of expected outcomes ("at Coffee Place A, I can get strong, flavourful coffee") to prospectively evaluate the current value of an action. Another key property of model-based evaluation is its sensitivity to changes in subjective value when a predicted outcome becomes more or less desirable ("going to Coffee Place A is particularly good if I'm tired, but is less good when I've recently had a coffee and am feeling energetic"). Thus, as model-free evaluation cannot yield behaviours that reveal knowledge of specific outcome information or exhibit sensitivity to changes in outcome value, Pavlovian conditioned responses that demonstrate such properties provide evidence of underlying model-based evaluations.

In their study, Pool and colleagues tested whether measures of Pavlovian learning might differ not only in the degree to which they reflect specific perceptual properties of an outcome but also in their sensitivity to the devaluation of that outcome[5]. Hungry participants completed a task in which two distinct visual cues probabilistically predicted the side of the screen, either left or right, on which a video 'reward' displaying the delivery of a favourite snack appeared. Two measures—pupil diameter and eye gaze—provided evidence of Pavlovian learning. Participants' pupil diameters were greater following cues that predicted reward, irrespective of the signalled laterality of the video, than following a third cue that did not predict a video, suggesting that this response reflected value information. During reward anticipation, participants also spent longer looking at the spatial location in which reward delivery was most probable, providing evidence of a learned association that incorporated specific features of the expected outcome. Control experiments demonstrated that eye gaze indeed represented a reward-sensitive Pavlovian response (i.e., participants looked longer at predicted locations when rewarding videos were expected than for matched videos with no food reward) and not an instrumental learned behaviour (i.e., participants directed their gaze to the video even when rewarded for looking away). After this learning phase, participants were allowed to consume as much of the favourite food that appeared in the videos as they desired, reducing the subjective value of this outcome. Following this opportunity to eat to satiety, participants again saw the predictive cues, enabling assessment of the degree to which each conditioned response was altered by the devaluation of the associated outcome.

Participants' pupil diameters during reward anticipation were smaller following outcome devaluation, providing evidence that this conditioned response adapted to the decrease in the subjective value of the outcome. Control experiments demonstrated that this devaluation-induced change in pupil dilation did not stem from overall decreases in reward motivation or generalized effects of satiety (i.e., pupil diameter decreased more for the devalued snack compared to another favourite snack that was not eaten). In contrast, the amount of time spent looking at the spatial location of the expected outcome did not show such sensitivity to devaluation. Intriguingly, both conditioned response measures—pupil diameter and eye gaze—showed behavioural patterns that cannot be accounted for by model-free computations. Sensitivity to changes in devaluation, evident in the pupil-diameter response, is a hallmark of model-based learning. Moreover, the eye-gaze response reflects knowledge of the expected spatial location of the video, an outcome feature that reveals an underlying model-based representation. However, the devaluation insensitivity evident in participants' eye-gaze responses has conventionally been thought to indicate underlying model-free computations, which require direct experience with an outcome that is no longer desirable in order to alter value estimates.

The persistence of this gaze response despite devaluation of the associated outcome presents a puzzle, as it reveals a learning mechanism that constructs a mental model of the features of a rewarding outcome but does not readily adapt to changes in that outcome's reward value. Such devaluation-insensitive attentional prioritization might reflect the retention of knowledge derived from Pavlovian learning (i.e., the location or sensory properties of a reward or threat) that might be adaptively leveraged to inform future action selection[2]. These provocative findings suggest that even a taxonomy distinguishing model-free and model-based forms of Pavlovian learning may be too simplistic to capture the diverse value computations that determine how we react to salient cues in our environment. ❐

Hillary A. Raab[1] and Catherine A. Hartley[1,2]*
[1]*Department of Psychology, New York University, New York, NY, USA.* [2]*Center for Neural Science, New York University, New York, NY, USA.*
**e-mail: cate@nyu.edu*

Published online: 25 February 2019
https://doi.org/10.1038/s41562-019-0538-1

References

1. Schultz, W., Dayan, P. & Montague, P. R. *Science* **275**, 1593–1599 (1997).
2. Robinson, M. J. & Berridge, K. C. *Curr. Biol.* **23**, 282–289 (2013).
3. Zhang, S., Mano, H., Ganesh, G., Robbins, T. & Seymour, B. *Curr. Biol.* **26**, 52–58 (2016).
4. Dayan, P. & Berridge, K. C. *Cogn. Affect. Behav. Neurosci.* **14**, 473–492 (2014).
5. Pool, E. R., Pauli, W. M., Kress, C. S. & O'Doherty, J. P. *Nat. Hum. Behav.* https://doi.org/10.1038/s41562-018-0527-9 (2018).
6. Daw, N. D., Niv, Y. & Dayan, P. *Nat. Neurosci.* **8**, 1704–1711 (2005).

Competing interests

The authors declare no competing interests.