Higher-Order Conditioning With Simultaneous and Backward Conditioned Stimulus: Implications for Models of Pavlovian Conditioning

Arthur Prevel, Ruth M. Krebs

Check for updates

Higher-Order Conditioning With Simultaneous and Backward Conditioned Stimulus: Implications for Models of Pavlovian Conditioning

Arthur Prével* and Ruth M. Krebs

Department of Experimental Psychology, Ghent University, Ghent, Belgium

In a new environment, humans and animals can detect and learn that cues predict meaningful outcomes, and use this information to adapt their responses. This process is termed Pavlovian conditioning. Pavlovian conditioning is also observed for stimuli that predict outcome-associated cues; a second type of conditioning is termed higher-order Pavlovian conditioning. In this review, we will focus on higher-order conditioning studies with simultaneous and backward conditioned stimuli. We will examine how the results from these experiments pose a challenge to models of Pavlovian conditioning like the Temporal Difference (TD) models, in which learning is mainly driven by reward prediction errors. Contrasting with this view, the results suggest that humans and animals can form complex representations of the (temporal) structure of the task, and use this information to guide behavior, which seems consistent with model-based reinforcement learning. Future investigations involving these procedures could result in important new insights on the mechanisms that underlie Pavlovian conditioning.

Keywords: backward conditioning, higher-order conditioning, reinforcement learning, reward prediction error, simultaneous conditioning

INTRODUCTION

When being exposed to a new environment, humans and other animals can detect and learn that cues or contextual stimuli predict the prospect of meaningful events. This learning process and the behavioral change associated are classically named Pavlovian conditioning (Hollis, 1997; Fanselow and Wassum, 2015). Not limited to the pairing between a stimulus and an outcome, Pavlovian conditioning is also observed for stimuli that predict outcome-associated cues. This second type of conditioning, in which a cue predicts another predictive stimulus, is referred to as higher-order Pavlovian conditioning (Gewirtz and Davis, 2000). Higher-order conditioning is particularly interesting as it is an excellent way to understand how humans and other animals form complex representations of the structure of the environment, and how they use these representations to guide flexible responses (Jones et al., 2012; Sadacca et al., 2016, 2018; Wang et al., 2020; Chandran and Thorwart, 2021). In the lab, higher-order conditioning is studied by second-order conditioning or sensory preconditioning (e.g., Gewirtz and Davis, 2000; Parkes and Westbrook, 2011). In second-order conditioning, a stimulus (CS1) is first paired with an unconditioned stimulus (US) until CS1 evokes a conditioned response (CR). Then, in a subsequent phase, a second stimulus (CS2) is paired with CS1 but without the US. At the end of the second phase, and despite the absence

of direct pairing with the US, the presentation of CS2 alone is sufficient to evoke a CR (see **Figures 1A–C**; Rizley and Rescorla, 1972; Rashotte et al., 1977). The pairing procedure used in sensory preconditioning is similar to second-order conditioning except that the order of phases 1 and 2 is inversed (i.e., CS2 → CS1 pairings, then CS1 → US pairings; Rescorla and Cunningham, 1978).

Traditionally, investigations on higher-order conditioning involve forward CS2 → CS1 and CS1 → US pairings. However, far less investigated are procedures involving simultaneous or backward pairings (e.g., Prével et al., 2019). In this mini-review, we will argue that these procedures are actually particularly relevant for the understanding of Pavlovian conditioning. Results from these experiments are indeed difficult to interpret in terms of the Reward Prediction Error (RPE) hypothesis (Schultz and Dickinson, 2000) and for models that implement this learning-rule like Temporal Difference (TD) learning models (Sutton and Barto, 2018). On the opposite end, the results seem to be conceptually consistent with model-based reinforcement learning systems (Daw et al., 2005; Gläscher et al., 2010; O'Doherty et al., 2017) and call for new investigations on the underlying computational mechanisms. After a presentation of the RPE hypothesis and a description of how a TD approach can account for higher-order conditioning, we will present results from higher-order conditioning studies that used simultaneous and backward pairing. We will discuss how far they are difficult to interpret from a reward prediction error perspective and how they seem to support model-based reinforcement learning systems. We will conclude this mini-review by discussing the perspectives offered by follow-up studies on higher-order conditioning with simultaneous and backward pairing.

REWARD PREDICTION ERROR AND HIGHER-ORDER CONDITIONING

Historically, one of the most dominant hypotheses about Pavlovian acquisition has been the RPE hypothesis (Schultz and Dickinson, 2000; Niv and Schoenbaum, 2008). This hypothesis states that a change in the value of a CS is driven by the discrepancy between the outcome expected from that stimulus, and the outcome actually received. Quantitative formulations of the RPE hypothesis are now largely based on TD learning (Niv and Schoenbaum, 2008; Ludvig et al., 2012; Sutton and Barto, 2018). Close to the Rescorla and Wagner (1972) model in terms of learning rule, TD models present the advantage of solving some of its important failures. The TD approach makes notably successful predictions about second-order conditioning, a phenomenon difficult to explain in terms of the Rescorla and Wagner model (Miller et al., 1995). In TD models, RPE (δ) is defined by:

$$\delta_{t+1} = R_{t+1} + \gamma V_{t+1} - V_t$$

Where R_{t+1} is the observed reward at $t+1$, V_{t+1} and V_t are the predicted value at $t+1$ and t, and γ is a discount

factor (with $0 < \gamma \leq 1$). δ is used to update the prediction made at t by:
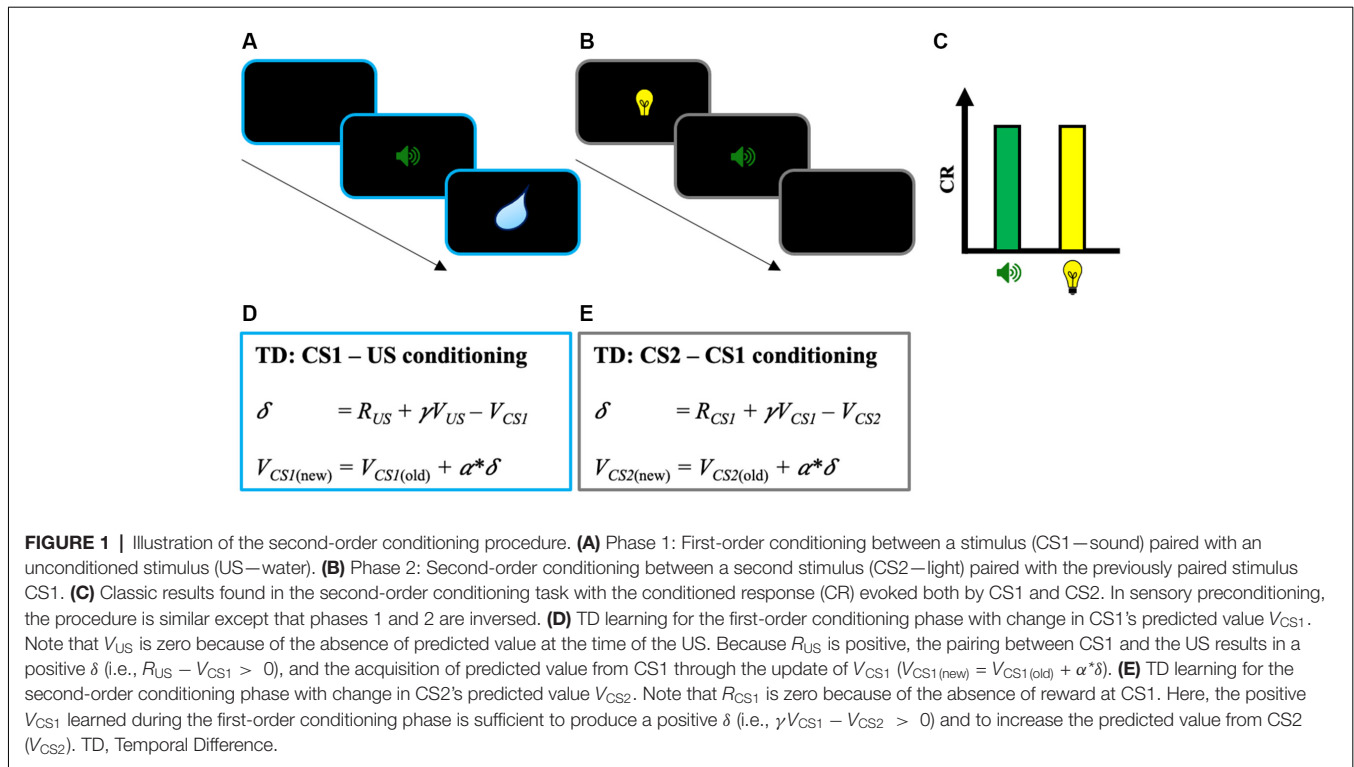
$$V_t = V_t + \alpha (\delta_{t+1})$$

Where α is a learning rate parameter (with $0 < \alpha \leq 1$).

Using this learning rule the TD models of Pavlovian conditioning can successfully explain second-order conditioning (Seymour et al., 2004; Sutton and Barto, 2018; Maes et al., 2020; see **Figures 1D,E**). In the first phase of the procedure, the pairing between CS1 and the US results in a positive δ and the acquisition of predicted value from CS1 (i.e., positive V_{CS1}). Then, this predicted value can be used to drive learning on CS2 in the second phase of the procedure. Despite the absence of reward during the second-order conditioning phase (i.e., $R_{CS1} = 0$), the positive value of V_{CS1} is sufficient to produce a positive δ (i.e., $\gamma V_{CS1} - V_{CS2} > 0$) and to increase the predicted value from CS2 (V_{CS2}). Interestingly, at the neural level, it has been found that the activity of dopaminergic neurons in a similar task moves backward from the US to the first predictive stimulus cue (i.e., CS2), as it would be predicted by TD models (Schultz, 2015).

Thus, TD learning seems particularly relevant to understanding the acquisition of higher-order predictive values, both at a behavioral and a neural level. The approach, however, is not without limitations. Particularly, the model fails to explain the acquisition of predictive value by CS2 in sensory preconditioning tasks: Due to the absence of reward in phase 1 and the predicted value of zero for CS1 (i.e., $R_{CS1} + \gamma V_{CS1} = 0$), a change in V_{CS2} is not expected according to TD models. However, the evidence from measuring responses to CS2 suggests the acquisition of predicted value from the stimulus. This challenge to TD learning has been repeatedly highlighted in the literature, and it becomes one of the arguments against the hypothesis that Pavlovian conditioning is only driven by RPE (Niv and Schoenbaum, 2008; Sadacca et al., 2016). Much less considered is the challenge posed by results from higher-order conditioning studies that involve a simultaneous or a backward CS1. Here, we believe that these results are particularly relevant for our understanding of higher-order learning. The next section will be dedicated to these findings.

HIGHER-ORDER CONDITIONING WITH SIMULTANEOUS AND BACKWARD PAIRING

In Pavlovian conditioning, the classic pairing procedure used to study the acquisition of new stimulus-outcome associations is the forward procedure in which the CS precedes the presentation of the US. Contrasting with this, in simultaneous and backward pairing the CS is presented simultaneously and after the US, respectively (**Figures 2A,B**). Experiments that used these procedures classically showed low response rates to the CS, or even the development of conditioned inhibition (Spooner and Kellogg, 1947; Fitzwater and Reisman, 1952; Moscovitch and LoLordo, 1968; Siegel and Domjan, 1974; but see

FIGURE 1 | Illustration of the second-order conditioning procedure. **(A)** Phase 1: First-order conditioning between a stimulus (CS1—sound) paired with an unconditioned stimulus (US—water). **(B)** Phase 2: Second-order conditioning between a second stimulus (CS2—light) paired with the previously paired stimulus CS1. **(C)** Classic results found in the second-order conditioning task with the conditioned response (CR) evoked both by CS1 and CS2. In sensory preconditioning, the procedure is similar except that phases 1 and 2 are inversed. **(D)** TD learning for the first-order conditioning phase with change in CS1's predicted value V_{CS1}. Note that V_{US} is zero because of the absence of predicted value at the time of the US. Because R_{US} is positive, the pairing between CS1 and the US results in a positive δ (i.e., $R_{US} - V_{CS1} > 0$), and the acquisition of predicted value from CS1 through the update of V_{CS1} ($V_{CS1(new)} = V_{CS1(old)} + \alpha * \delta$). **(E)** TD learning for the second-order conditioning phase with change in CS2's predicted value V_{CS2}. Note that R_{CS1} is zero because of the absence of reward at CS1. Here, the positive V_{CS1} learned during the first-order conditioning phase is sufficient to produce a positive δ (i.e., $\gamma V_{CS1} - V_{CS2} > 0$) and to increase the predicted value from CS2 (V_{CS2}). TD, Temporal Difference.

Spetch et al., 1981; Prével et al., 2016). These observations suggest that simultaneous and backward pairings are not appropriate procedures to produce a robust CR, which is consistent with TD models: When a simultaneous or a backward CS is presented, the stimulus is never followed by a reward at $t + 1$. Thus, a change in V_{CS} is consequently not expected from those pairing procedures. In addition, a higher-order cue (CS2) that precedes a simultaneous or a backward CS1 should not produce robust responding because V_{CS1} is zero at the end of the first-order conditioning phase.

From a functional perspective, the absence of a robust CR in simultaneous and backward pairing is not surprising if we consider that the function of the response is to prepare the organism for the US (Hollis, 1997). Because the CS is not predictive of the US, there is a priori no reason to expect a preparatory response evoked by that stimulus. However, what is not clear is whether the absence of a CR measured to the simultaneous or backward CS really means that subjects did not learn anything from these pairing procedures due to the RPE of zero. Alternatively, it is possible that subjects in these experiments learned an association between the simultaneous or backward CS and the US, but these associations are simply not overtly expressed due to the absence of predictive value of the CS (Arcediano and Miller, 2002). In what follows, we will discuss the results from higher-order conditioning studies that support this interpretation.

For example, Barnet et al. (1991) tested whether a first-order stimulus CS1 paired simultaneously with a US can support the conditioning of a second-order stimulus CS2 (see **Figure 2C** for an illustration). Consistent with common findings

in simultaneous conditioning studies, the authors reported low responses evoked by CS1 in comparison to a forward first-order stimulus, supporting the idea that the procedure is not efficient to produce a robust CR. However, when in a subsequent phase a second-order stimulus CS2 was paired with CS1 using a forward pairing (i.e., CS2 → CS1 pairings), the authors found a substantial level of CR evoked by CS2, despite the low response measured on CS1. These results by Barnet et al. (1991) seem difficult to explain in terms of TD learning. According to the account described above, a change in CS2 value (V_{CS2}) depends directly on CS1's own value (V_{CS1}). Thus, a second-order pairing with a first-order stimulus CS1 that evokes low response (and with presumably a low predicted value) should result in low response to CS2. However, the evidence of substantial response to that stimulus challenges this interpretation. Later, Barnet and Miller (1996) extended their investigations to backward conditioning. In phase 1 of a second-order conditioning task, a first-order stimulus CS1 was paired to a US using backward pairing. This resulted in the development of conditioned inhibition, a classic result of this pairing procedure. Interestingly, when in phase 2 a second-order stimulus CS2 was paired with CS1 using forward pairing, this resulted in substantial CR to CS2 despite the inhibitory status of CS1. Again, the result is problematic for the TD account of second-order conditioning. It is not clear why a first-order stimulus CS1 with an acquired inhibitory status (and presumably, a negative predicted value V_{CS1}) can support the conditioning of a second-order stimulus CS2.

These results by Barnet and Miller (1996) were replicated by Cole and Miller (1999), who found that the effect varied
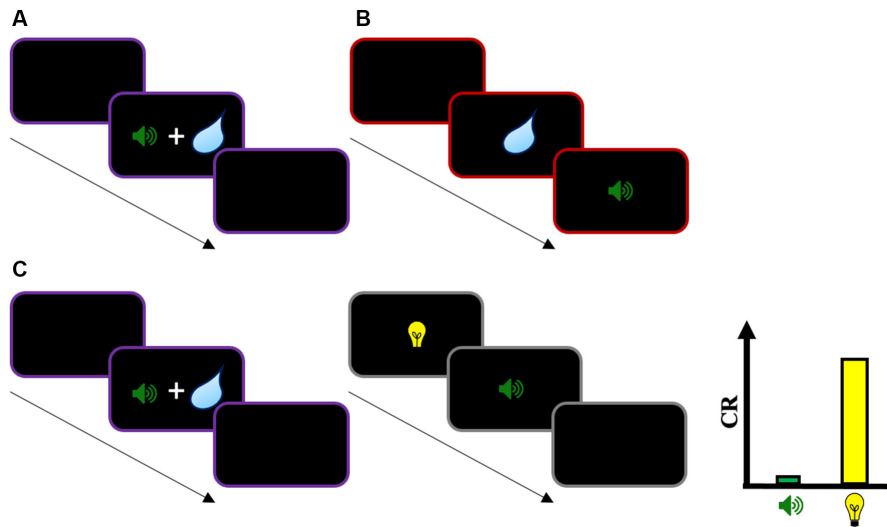
FIGURE 2 | Illustration of simultaneous, backward, and second-order conditioning with simultaneous CS1. **(A)** Simultaneous conditioning with a stimulus (CS1—sound) presented simultaneously with an unconditioned stimulus (US—water). **(B)** Backward conditioning with CS1 presented after a US. **(C)** Second-order conditioning with simultaneous CS1: In phase 1, a stimulus CS1 is presented simultaneously with a US. In phase 2, a second stimulus CS2 is paired with CS1 through forward pairing. During the test, while CS1 will evoke low conditioned response (CR), CS2 will evoke substantial CR. According to the TD account, a low CR evoked by CS1 is expected because CS1 is not followed by the US (i.e., $R_{US} = 0$) in phase 1. In addition, a change in CS2 value (V_{CS2}) depends directly on CS1's own value (V_{CS1}). Thus, a second-order pairing with a first-order stimulus CS1 that evokes a low CR level (and with presumably a low predicted value) should result in low responding to CS2. The evidence of substantial response to that stimulus challenges the TD account. The same holds for a model-based account of higher-order learning if the change in V_{CS2} depends on CS1's own predicted value V_{CS1}. Instead, it seems necessary for CS2's predicted value to be based on US expectations to account for this finding. Note that the same pattern of results is observed for second-order conditioning with backward CS1, and for sensory preconditioning with simultaneous and backward CS1.

with the number of backward pairing trials in phase 1. More exactly, the authors reported that a backward CS1 supports second-order conditioning only when the number of backward pairing trials is low or high, but the CR to CS2 decreases at an intermediate number of trials. Parallel to these investigations, Barnet et al. (1997) demonstrated that a backward CS1 can support stronger second-order conditioning compared to a forward first-order CS1, despite a lower CR to that backward stimulus. More recent observations by Prével et al. (2019) are consistent with these findings. Specifically, the authors demonstrated that a second-order stimulus CS2 can function as an efficient conditioned reinforcer for instrumental response in the test phase, even when that stimulus was paired with a backward CS1 that did not evoke CR during phase 2. Finally, similar findings were reported using sensory preconditioning. For example, Matzel et al. (1988) found evidence of substantial sensory preconditioning with simultaneous and backward first-order paired stimuli. Barnet et al. (1997) reported results similar to their observations in a second-order conditioning task with sensory preconditioning. Finally, Arcediano et al. (2003) found successful sensory preconditioning with backward first-order CS. In summary, it seems clear from all these experiments that a simultaneous or backward first-order CS can support higher-order learning, even if that same stimulus shows a low CR level or conditioned inhibition. As we have seen, the evidence is difficult to explain based on TD models, and particularly with regard to sensory preconditioning due to the additional absence of RPE

in phase 1. In the next section, we will describe the model-based reinforcement learning account as a valuable alternative to TD learning.

MODEL-BASED REINFORCEMENT LEARNING AND HIGHER-ORDER CONDITIONING

Because of the challenges posed by effects like sensory preconditioning, the last 10–20 years have seen the development of another class of models termed model-based reinforcement learning (Daw et al., 2005; Gläscher et al., 2010; O'Doherty et al., 2017). In this approach, human subjects and animals can learn a model of the environment to guide appropriate responding. This model includes the states encountered by the subjects, as well as the transition probabilities between states and the available rewards. This contrasts with (model-free) TD models in which the subjects merely learn the predicted value of each state, but not the potential transition between states. Another characteristic of the model-based approach resides in the fact that the subjects can use the learned-transitions between states to update the states' value through a (mental) simulation mechanism. This second aspect is particularly interesting because it can be used to account for goal-directed phenomena like devaluation (e.g., Wilson et al., 2014), but certainly also sensory preconditioning: Here, during phase 1 participants would learn the transition probability between CS2 and CS1, before learning during phase

2 the positive predictive value of CS1 based on its direct pairing with the US. Then, through a simulation mechanism, the learned transition between CS2 and CS1 and the expected value from V_{CS1} could be used to update V_{CS2}, i.e., subjects could (mentally) assign a new value to CS2 based on the learned-transition between CS2 and CS1 (i.e., CS2 is followed by CS1), and the learned predicted value from CS1. For example, if we adapt the model-based mechanism proposed by Wilson et al. (2014) to sensory preconditioning, at the end of training the (model-based) value of CS2 could be updated through:

$$V_{CS2} = V_{CS1} \times p \, (CS1 \parallel CS2)$$

Where $p(CS1 \parallel CS2)$ is the estimated learned probability of CS2 leading to CS1, and V_{CS1} is the predicted value from CS1. Because $p(CS1 \parallel CS2)$ and V_{CS1} are positive due to the pairings in phases 1 and 2, this would result in a positive V_{CS2} and the ability of the stimulus to evoke CR.

In addition to sensory preconditioning, the model-based learning approach seems also very promising to account for the findings presented in the previous section. The assumption that humans and animals can learn a model representing the structure of the environment, and that they use this model to flexibly update the value of states (stimuli) and guide responding, seems remarkably consistent with the results described above. In these experiments, it is as if subjects learned the (temporal) structure of the task and used this structure to infer a predictive value from CS2 and guide responding: Participants first learned that CS1 is presented simultaneously or after the US, but the absence of predictive value of CS1 prevented the development of a robust CR. However, through the integration of the associations learned in phases 1 and 2, the forward pairing between CS2 and CS1 conferred a predictive value between CS2 and (the representation of) the US, which resulted in the CR measured in response to this stimulus (see Arcediano and Miller, 2002). Interestingly, multiple results in the literature suggest the acquisition of such temporal maps (e.g., Cole et al., 1995; Arcediano et al., 2005; Thrailkill and Shahan, 2014). However, it must be noted that it is not clear what the exact computational mechanism is that supports the temporal integration and the acquired predicted value on CS2 observed in these studies. If we consider for example the model-based mechanism described above, because a change in V_{CS2} depends in this formulation on CS1's own predicted value V_{CS1}, the problem remains that it is difficult to understand why a stimulus that shows low CR or conditioned inhibition supports substantial CR to CS2. Instead, it seems necessary for CS2's predicted value to be based on US expectation to explain the results presented in the previous section. More investigations will be necessary to propose a complete account of higher-order learning, and particularly a mechanism that allows the temporal integration of the task structure to guide flexible and adaptive responses.

CONCLUSION AND OUTLOOK

The evidence from sensory preconditioning and higher-order conditioning with simultaneous and backward pairing pose a challenge to the assumption that Pavlovian conditioning is driven only by RPE. Rather, these observations suggest that subjects were able to learn a representation of the (temporal) structure of the task and to use this representation to guide their responses, which seems consistent with the assumptions of model-based reinforcement learning. However, the exact nature of the computational mechanisms is still missing. Here, we are highlighting three fruitful directions for future investigations on higher-order conditioning with simultaneous and backward CS. First, it must be noted that model-free reinforcement learning approaches such as TD models are not necessarily dismissed by these results. To the best of our knowledge, the consensus in the literature seems to assume a co-existence of both model-free and model-based reinforcement learning systems, representing habitual and goal-directed behaviors, respectively (Gläscher et al., 2010; Wilson et al., 2014; O'Doherty et al., 2017). Additional investigations on higher-order conditioning with simultaneous and backward pairing could provide new insights regarding the computational mechanisms that underly model-based reinforcement learning and temporal integration in higher-order conditioning, as well as how model-free and model-based reinforcement learning computations are integrated in that context. Second, an important research question in the study of higher-order conditioning concerns the nature of the associations learned (Gewirtz and Davis, 2000). New investigations using the procedures described in this mini-review could give new insights into what is learned by subjects in these tasks, which in turn could have important implications on the underlying computational mechanisms. Finally, an important hypothesis in the neuroscientific domain is that the phasic activity of dopaminergic neurons represents the RPE teaching signal in the context of model-free reinforcement learning (Schultz, 2015). However, recent results suggest instead that this activity could reflect model-based computations (Sadacca et al., 2016; Sharpe et al., 2017; Langdon et al., 2018; Sharpe and Schoenbaum, 2018). Here, it might be interesting to study how this activity changes during the presentation of CS1 and CS2 depending on the pairing conditions and to test which neural structures are subserving task representations and value updates.

AUTHOR CONTRIBUTIONS

Both authors contributed to the article and approved the submitted version.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Arcediano, F., Escobar, M., and Miller, R. R. (2003). Temporal integration and temporal backward associations in human and nonhuman subjects. *Learn. Behav.* 31, 242–256. doi: 10.3758/bf03195986

Arcediano, F., Escobar, M., and Miller, R. R. (2005). Bidirectional associations in humans and rats. *J. Exp. Psychol. Anim. Behav. Process.* 31, 301–318. doi: 10.1037/0097-7403.31.3.301

Arcediano, F., and Miller, R. R. (2002). Some constraints for models of timing: a temporal coding hypothesis perspective. *Learn. Motiv.* 33, 105–123. doi: 10.1006/lmot.2001.1102

Barnet, R. C., Arnold, H. M., and Miller, R. R. (1991). Simultaneous conditioning demonstrated in second-order conditioning: evidence for similar associative structure in forward and simultaneous conditioning. *Learn. Motiv.* 22, 253–268. doi: 10.1016/0023-9690(91)90008-V

Barnet, R. C., Cole, R. P., and Miller, R. R. (1997). Temporal integration in second-order conditioning and sensory preconditioning. *Anim. Learn. Behav.* 25, 221–233. doi: 10.3758/BF03199061

Barnet, R. C., and Miller, R. R. (1996). Second-order excitation mediated by a backward conditioned inhibitor. *J. Exp. Psychol. Anim. Behav. Process.* 22, 279–296. doi: 10.1037//0097-7403.22.3.279

Chandran, M., and Thorwart, A. (2021). Time in associative learning: a review on temporal maps. *Front. Hum. Neurosci.* 15:617943. doi: 10.3389/fnhum.2021.617943

Cole, R. P., Barnet, R. C., and Miller, R. R. (1995). Temporal encoding in trace conditioning. *Anim. Learn. Behav.* 23, 144–153. doi: 10.3758/BF03199929

Cole, R. P., and Miller, R. R. (1999). Conditioned excitation and conditioned inhibition acquired through backward conditioning. *Learn. Motiv.* 30, 129–156. doi: 10.1006/lmot.1998.1027

Daw, N. D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* 8, 1704–1711. doi: 10.1038/nn1560

Fanselow, M. S., and Wassum, K. M. (2015). The origins and organization of vertebrate pavlovian conditioning. *Cold Spring Harb. Perspect. Biol.* 8:a021717. doi: 10.1101/cshperspect.a021717

Fitzwater, M. E., and Reisman, M. N. (1952). Comparisons of forward, simultaneous, backward and pseudo-conditioning. *J. Exp. Psychol.* 44, 211–214. doi: 10.1037/h0060495

Gewirtz, J. C., and Davis, M. (2000). Using pavlovian higher-order conditioning paradigms to investigate the neural substrates of emotional learning and memory. *Learn. Mem.* 7, 257–266. doi: 10.1101/lm.35200

Gläscher, J., Daw, N., Dayan, P., and O'Doherty, J. P. (2010). States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron* 66, 585–595. doi: 10.1701/3679.36651

Hollis, K. L. (1997). Contemporary research on Pavlovian conditioning. A "new" functional analysis. *Am. Psychol.* 52, 956–965. doi: 10.1037//0003-066x.52.9.956

Jones, J. L., Esber, G. R., McDannald, M. A., Gruber, A. J., Hernandez, A., Mirenzi, A., et al. (2012). Orbitofrontal cortex supports behavior and learning using inferred but not cached values. *Science* 338, 953–956. doi: 10.1126/science.1227489

Langdon, A. J., Sharpe, M. J., Schoenbaum, G., and Niv, Y. (2018). Model-based predictions for dopamine. *Curr. Opin. Neurobiol.* 49, 1–7. doi: 10.1016/j.conb.2017.10.006

Ludvig, E. A., Sutton, R. S., and Kehoe, E. J. (2012). Evaluating the TD model of classical conditioning. *Learn. Behav.* 40, 305–319. doi: 10.3758/s13420-012-0082-6

Maes, E. J., Sharpe, M. J., Usypchuk, A. A., Lozzi, M., Chang, C. Y., Gardner, M. P., et al. (2020). Causal evidence supporting the proposal that dopamine transients function as temporal difference prediction errors. *Nat. Neurosci.* 23, 176–178. doi: 10.1038/s41593-019-0574-1

Matzel, L. D., Held, F. P., and Miller, R. R. (1988). Information and expression of simultaneous and backward associations: implications for contiguity theory. *Learn. Motiv.* 19, 317–344. doi: 10.1016/0023-9690(88)90044-6

Miller, R. R., Barnet, R. C., and Grahame, N. J. (1995). Assessment of the Rescorla-Wagner model. *Psychol. Bull.* 117, 363–386. doi: 10.1037/0033-2909.117.3.363

Moscovitch, A., and LoLordo, V. M. (1968). Role of safety in the Pavlovian backward fear conditioning procedure. *J. Comp. Physiol. Psychol.* 66, 673–678. doi: 10.1037/h0026548

Niv, Y., and Schoenbaum, G. (2008). Dialogues on prediction errors. *Trends Cogn. Sci.* 12, 265–272. doi: 10.1016/j.tics.2008.03.006

O'Doherty, J. P., Cockburn, J., and Pauli, W. M. (2017). Learning, reward and decision making. *Annu. Rev. Psychol.* 68, 73–100. doi: 10.1146/annurev-psych-010416-044216

Parkes, S. L., and Westbrook, R. F. (2011). Role of the basolateral amygdala and NMDA receptors in higher-order conditioned fear. *Rev. Neurosci.* 22, 317–333. doi: 10.1515/RNS.2011.025

Prével, A., Rivière, V., Darcheville, J.-C., and Urcelay, G. P. (2016). Conditioned reinforcement and backward association. *Learn. Motiv.* 56, 38–47. doi: 10.1016/j.lmot.2016.09.004

Prével, A., Rivière, V., Darcheville, J. C., Urcelay, G. P., and Miller, R. R. (2019). Excitatory second-order conditioning using a backward first-order conditioned stimulus: a challenge for prediction error reduction. *Q. J. Exp. Psychol. (Hove)* 72, 1453–1465. doi: 10.1177/1747021818793376

Rashotte, M. E., Griffin, R. W., and Sisk, C. L. (1977). Second-order conditioning of the pigeon's keypeck. *Anim. Learn. Behav.* 5, 25–38.

Rescorla, R. A., and Cunningham, C. L. (1978). Within-compound flavor associations. *J. Exp. Psychol. Anim. Behav. Process.* 4, 267–275. doi: 10.1037/0097-7403.4.3.267

Rescorla, R. A., and Wagner, A. R. (1972). "A theory of Pavlovian conditioning: variations on the effectiveness of reinforcement and non-reinforcement," in *Classical conditioning II: Current Research and Theory*, eds A. H. Black, and W. F. Prokasy (New York: Appleton-Century-Crofts), 64–99.

Rizley, R. C., and Rescorla, R. A. (1972). Associations in second-order conditioning and sensory preconditioning. *J. Comp. Physiol. Psychol.* 81, 1–11. doi: 10.1037/h0033333

Sadacca, B. F., Jones, J. L., and Schoenbaum, G. (2016). Midbrain dopamine neurons compute inferred and cached value prediction errors in a common framework. *eLife* 5:e13665. doi: 10.7554/eLife.13665

Sadacca, B. F., Wied, H. M., Lopatina, N., Saini, G. K., Nemirovsky, D., and Schoenbaum, G. (2018). Orbitofrontal neurons signal sensory associations underlying model-based inference in a sensory preconditioning task. *eLife* 7:e30373. doi: 10.7554/eLife.30373

Schultz, W. (2015). Neuronal reward and decision signals: from theories to data. *Physiol. Rev.* 95, 853–951. doi: 10.1152/physrev.00023.2014

Schultz, W., and Dickinson, A. (2000). Neuronal coding of prediction errors. *Annu. Rev. Neurosci.* 23, 473–500. doi: 10.1146/annurev.neuro.23.1.473

Seymour, B., O'Doherty, J. P., Dayan, P., Koltzenburg, M., Jones, A. K., Dolan, R. J., et al. (2004). Temporal difference models describe higher-order learning in humans. *Nature* 429, 664–667. doi: 10.1038/nature02581

Sharpe, M. J., Chang, C. Y., Liu, M. A., Batchelor, H. M., Mueller, L. E., Jones, J. L., et al. (2017). Dopamine transients are sufficient and necessary for acquisition of model-based associations. *Nat. Neurosci.* 20, 735–742. doi: 10.1038/nn.4538

Sharpe, M. J., and Schoenbaum, G. (2018). Evaluation of the hypothesis that phasic dopamine constitutes a cached-value signal. *Neurobiol. Learn. Mem.* 153, 131–136. doi: 10.1016/j.nlm.2017.12.002

Siegel, S., and Domjan, M. (1974). The inhibitory effect of backward conditioning as a function of the number of backward pairings. *Bull. Psychon. Soc.* 4, 122–124. doi: 10.3758/BF03334216

Spetch, M. L., Wilkie, D. M., and Pinel, J. P. (1981). Backward conditioning: a reevaluation of the empirical evidence. *Psychol. Bull.* 89, 163–175.

Spooner, A., and Kellogg, W. N. (1947). The backward conditioning curve. *Am. J. Psychol.* 60, 321–334.

Sutton, R. S., and Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. Cambridge, MA: The MIT Press.

Thrailkill, E. A., and Shahan, T. A. (2014). Temporal integration and instrumental conditioned reinforcement. *Learn. Behav.* 42, 201–208. doi: 10.3758/s13420-014-0138-x

Wang, F., Schoenbaum, G., and Kahnt, T. (2020). Interactions between human orbitofrontal cortex and hippocampus support model-based inference. *PLoS Biol.* 18:e3000578. doi: 10.1371/journal.pbio.3000578

Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., and Niv, Y. (2014). Orbitofrontal cortex as a cognitive map of task space. *Neuron* 81, 267–279. doi: 10.1016/j.neuron.2013.11.005