Mesolimbic dopamine release conveys causal associations

Huijeong Jeong[1], Annie Taylor[2]†, Joseph R. Floeder[2]†, Martin Lohmann[4], Stefan Mihalas[4,5], Brenda Wu[1], Mingkang Zhou[1,2], Dennis A. Burke[1], Vijay Mohan K Namboodiri[1,2,3]*

[1]Department of Neurology, University of California, San Francisco, CA, USA. [2]Neuroscience Graduate Program, University of California, San Francisco, CA, USA. [3]Weill Institute for Neuroscience, Kavli Institute for Fundamental Neuroscience, Center for Integrative Neuroscience, University of California, San Francisco, CA, USA. [4]Allen Institute for Brain Science, Seattle, WA, USA. [5]Department of Applied Mathematics, University of Washington, Seattle, WA, USA.

*Corresponding author. E-mail: VijayMohan.KNamboodiri@ucsf.edu

Learning to predict rewards based on environmental cues is essential for survival. It is believed that animals learn to predict rewards by updating predictions whenever the outcome deviates from expectations, and that such reward prediction errors (RPEs) are signaled by the mesolimbic dopamine system—a key controller of learning. However, instead of learning prospective predictions from RPEs, animals can infer predictions by learning the retrospective cause of rewards. Hence, whether mesolimbic dopamine instead conveys a causal associative signal that sometimes resembles RPE remains unknown. We developed an algorithm for retrospective causal learning and found that mesolimbic dopamine release conveys causal associations but not RPE, thereby challenging the dominant theory of reward learning. Our results reshape the conceptual and biological framework for associative learning.

How do animals learn to associate environmental cues with delayed outcomes such as rewards? It is widely believed that they learn a prospective prediction of how often reward follows a given cue. A simple way to learn such prospective predictions is to update one's prediction every time the outcome following a cue deviates from the prediction (Fig. 1, A and B). Such violations of reward predictions are commonly called reward prediction errors (RPEs). The simplest model in this family is the Rescorla-Wagner model (*1*). Temporal difference reinforcement learning (TDRL) models extend the Rescorla-Wagner model to account for cue-outcome delays and is the most widely accepted model of reward learning (*2*, *3*). To account for delays, these models typically propose that a sequential pattern of neural activities ("states") tiles temporal delays and propagates predictions from the cue to the reward (Fig. 1B). TDRL RPE has been successful at explaining the activity dynamics of dopaminergic cell bodies and release in the nucleus accumbens (*4–13*). Hence, TDRL RPE has become the dominant theory of dopamine's role as the critical regulator of behavioral learning (*14–17*).

An alternative approach to learn cue-reward associations is to infer the cause of meaningful outcomes such as rewards (*18–20*) (Fig. 1, A and C). Because causes must precede outcomes, a viable approach to infer whether a cue causes a reward is to learn whether the cue consistently precedes a reward. Predicting the future is highly demanding in a cue-rich environment but inferring the cause of a rarer meaningful outcome simply requires a memory of previous experience. If an animal knows that a stimulus it just received is meaningful (e.g., a reward), it can look back in memory to infer its cause. Given the central role of dopamine in learning, we hypothesized that dopamine may guide retrospective causal learning instead of conveying RPEs. Though the

differences between prospective and retrospective learning may not be apparent at first glance, we show that these models make highly divergent predictions about mesolimbic dopamine dynamics. Here, we directly test between these models of the role of dopamine in associative learning.

A retrospective causal learning algorithm

While some stimuli are innately meaningful others acquire meaning after learning that they cause other meaningful stimuli (e.g., a cue that predicts a reward becomes meaningful). We denote stimuli whose cause should be learned by the animal as a "meaningful causal targets" and propose that mesolimbic dopamine signals whether a current event is a meaningful causal target (Fig. 1C and figs. S1 and S2). We propose a causal inference algorithm that infers whether a cue is a cause of reward by measuring whether it precedes the reward more than that expected by chance (Fig. 1C and fig. S2), then converting this to a prospective prediction signal using Bayes' rule (fig. S3) (Supplementary Note 1), and finally using the net contingency between a cue and reward to build a cognitive map of causal associations (*20*) (Fig. 1C and fig. S4).

We developed this algorithm to address problematic temporal assumptions that are foundational to common conceptions of TDRL, which result in a non-scalable representation of time (*21*). We tested whether this new algorithm learns causal relationships without loss of generality across timescales. Consistent with this and unlike TDRL, our algorithm learns the underlying causal structure of a variety of complex environments across two orders of magnitude of timescales and explains well-established behavioral observations of the timescale invariance of learning (figs. S5 and S6). The algorithm proposes that meaningful causal targets are signaled by an adjusted net contingency for causal

relations (ANCCR, read "anchor") () (fig. S4). The ANCCR-based causal learning model is consistent with simulations of classical results supporting the RPE coding hypothesis including dopaminergic responses to reward magnitude and probability, blocking, unblocking, overexpectation, conditioned inhibition, and trial-by-trial update of action probabilities (Fig. 2). It is also consistent with the observation that apparent negative RPEs in dopamine response are not as strong as positive RPEs of the same magnitude, even without assuming a floor effect in dopamine responses. Therefore, we reasoned that mesolimbic dopamine release has been tested only under conditions in which the ANCCR and RPE hypotheses make similar predictions, and that dopamine release may convey ANCCR instead of RPE.

In most behavioral tasks, prospective and retrospective associations are highly correlated and difficult to separate. To distinguish between the two hypotheses (RPE or ANCCR signaling by mesolimbic dopamine release), we performed eleven experimental tests. To maximize our ability to distinguish the models for strong inference (*22*), we designed the experiments such that the predictions of the two hypotheses are qualitatively different and often opposing. Because it has been proposed that distinct dopaminergic systems exist in the midbrain and that only some faithfully signal RPE (*23–30*), we tested these predictions by optically measuring sub-second mesolimbic dopamine release in the nucleus accumbens core (NAcc), a projection widely believed to encode RPE and shown to mediate Pavlovian learning (*8–10*, *12*, *31–33*) [by contrast, see (*34*)] (Fig. 3A and fig. S7). We did so in mice using fiber photometry of the dopamine sensor dLight 1.3b expressed in NAcc (*7*, *35*).

Tests 1 and 2 (unpredicted rewards)

We first tested between the two hypotheses in a simple experiment with divergent predictions. We presented naïve head-fixed mice with no experience in any laboratory behavior task with random unpredicted drops of a 15% sucrose solution delivered with an exponential inter-reward interval (IRI) distribution (mean = 12 s), while recording mesolimbic dopamine release in NAcc. In this task, the timing of individual sucrose deliveries cannot be anticipated based on the previous delivery, but the average rate of sucrose delivery is fixed (once every 12 s on average). Because the animal is experimentally naïve with no history of receiving sucrose prior to the onset of the experiment, the RPE hypothesis predicts high dopamine response to sucrose during the early exposures. This is because the sucrose is highly unpredicted initially. With repeated exposure to the context, the RPE is predicted to decrease slightly as the context becomes a predictor of the rewards. More formally, the internal IRI "states" in TDRL acquire positive value with experience (see Supplementary Note 2 for a consideration of a semi-Markov state space in TDRL) (*36*). Since RPE is the difference between the value of sucrose and the value of the IRI state that preceded sucrose delivery, RPE will reduce at sucrose delivery with repeated experience (Fig. 3, B and C).

On the other hand, the ANCCR hypothesis predicts that the response to sucrose will increase with repeated experience. This is because the predicted sucrose response is proportional to the difference between the average rate of previous sucrose deliveries calculated at sucrose delivery (including the current sucrose delivery) and the baseline average rate of previous sucrose deliveries (Fig. 3B). Because both of these quantities are initially low in naïve animals that have no experience with sucrose, ANCCR of sucrose is low early in this task. ANCCR eventually reaches an asymptote of ~1 times the incentive value of sucrose (Methods) because the rate of sucrose calculated just prior to a sucrose delivery (i.e., excluding the current sucrose) is equal to the baseline average rate of sucrose. Thus, the RPE hypothesis predicts that the dopamine response to sucrose will decrease over repeated experiences, while the ANCCR hypothesis predicts that the response will increase. Testing these differential predictions formed Test 1 (Fig. 3, B and C).

Observed mesolimbic dopamine release was consistent with ANCCR but not RPE (Fig. 3, D and E). Every animal showed an increasing sucrose response that reached a high positive asymptote. This is entirely inconsistent with RPE: because RPE is the difference between a received and predicted reward, it cannot be higher than that for an unpredicted reward. These results also cannot be explained by RPE based on a slower learning of the incentive value of sucrose; animals actively licked to consume sucrose at high rates starting from the first delivery, demonstrating that sucrose had high value (Fig. 3D and fig. S8). Such high motivation for sucrose from the onset of the experiment is consistent with well-known results that sugar is innately rewarding to mice (*37*). We also ruled out alternative hypotheses such as stress (Supplementary Note 3, fig. S8) or a non-specific increase in responses to the consummatory action (lick bout onset) (fig. S8).

We next tested a "trial-by-trial" prediction in this experiment by measuring the correlation between the dopamine response to a sucrose delivery and the previous IRI. Getting the next reward sooner than predicted would produce a larger RPE than getting the next reward later. Hence, the RPE hypothesis predicts a negative correlation between the dopamine response to a sucrose delivery and the previous IRI (*36*) (Fig. 3, B and F) (Supplementary Note 4). However, ANCCR predicts a positive correlation because the ANCCR of reward involves the subtraction of the baseline reward rate. Because the baseline reward rate declines with longer IRI, ANCCR should increase with longer IRI (Fig. 3, B and F). This was Test 2.

The experimentally observed correlation between dopamine response to sucrose and the previous IRI was positive, thereby being consistent with ANCCR but not RPE. We also ruled out the hypothesis that this positive correlation is simply due to an inability of animals to learn the mean IRI. This is because 1) the

correlation was consistently positive for more than 800 experiences of sucrose (8 sessions) (fig. S8), 2) mice learn the average IRI within at most two sessions (fig. S8), 3) rodents can be as fast as Bayesian ideal observers in detecting changes in the rate of exponentially scheduled rewards (*38*), and 4) even the original experiments that inspired the Rescorla-Wagner model showed that animals learn the mean inter-reinforcer interval despite unpredictable timing (*39*, *40*) (see (*41*) for a detailed discussion).

Tests 3 to 7 (Cue-reward learning)

Next, we studied dopamine response dynamics during cue-reward learning. We measured behavioral learning using anticipatory licking prior to the delivery of sucrose 3 s following onset of an auditory cue. Anticipatory licking reflects the prediction of upcoming reward across species, and this paradigm has provided some of the strongest support for TDRL RPE coding (*4*, *5*, *42–45*). During cue-reward learning, both RPE and ANCCR predict that dopamine responses to the cue will be low early in learning and high late in learning. Thus, the increase in dopamine response to cue can be used as a measure of dopaminergic learning (defined as dopaminergic signaling related to the external cue-reward association). The RPE hypothesis predicts a tight relationship between the dynamics of behavioral and dopaminergic learning (Fig. 4A). This is because TDRL RPE updates the value signal used for behavioral learning, and dopaminergic signaling in NAcc is necessary for the learning of anticipatory licking in head-fixed mice (*33*). On the other hand, the ANCCR of the cue is a continuously evolving estimate of whether the cue is itself a meaningful causal target due to its association with reward and hence is not predicted to evolve in lockstep with the behavior. Indeed, in the ANCCR hypothesis, associations are learned first, and then timing is learned: behavioral learning requires the threshold crossing of ANCCR to learn a causal model of the world ("cue causes reward"), followed by the separate learning of the temporal delay between cue and reward ("cue causes reward at a 3 s delay"). Only then does a timed decision signal for behavior become available (fig. 4B and fig. S2). Thus, the ANCCR hypothesis predicts that the gradual dopaminergic learning of the cue response will significantly precede behavioral learning, and that behavioral learning will be much more abrupt than dopaminergic learning since it requires an internal threshold crossing of the net contingency between cue and reward (Test 3) (Supplementary Note 5). The observed dopaminergic dynamics during learning were consistent with ANCCR but not RPE: dopamine response to CS+ was evident long before animals showed anticipatory licking (fig. 4, B and F, and fig. S9). In fact, dopamine cue responses were at their peak by the time of behavioral acquisition (fig. S10).

Further, when a learned delay between cue onset and reward (3 s) is extended permanently to a new, longer delay (9 s), RPE predicts that as animals learn the longer delay and suppress anticipatory licking at the previous short delay, there will be a concomitant reduction in the dopamine cue response due to temporal discounting (*46*). On the other hand, ANCCR predicts little to no change in the dopamine cue response as the structure of the task is largely unchanged (Test 4, Fig. 4G and figs. S9 and S10; intuitively, relative to the long intertrial interval, the cue-reward delay is still short). Experimentally, we observed that while the animals learned the new delay rapidly, dopaminergic cue response showed no significant change (Fig. 4, G to I). After the extension of the cue-reward delay, RPE predicts a suppression of dopamine after the old delay expires without reward. Because the increase in cue-reward delay is permanent [unlike in prior experiments (*45*)], ANCCR predicts that the delay representation in the internal causal model of the animal would be updated to reflect the new delay. This predicts no reward omission response at the old delay (3 s) after the increase in the delay to 9 s. Thus, ANCCR predicts no negative omission response after the old delay expires without reward. (Test 5). Experimentally, we observed no suppression of dopamine response at 3 s in this experiment but did observe suppression in a separate experiment when the reward was indeed omitted (Fig. 4J and fig. S10).

Next, we tested extinction of a learned cue-reward association. Extinction of a learned association does not cause unlearning of the original association (*47*). Yet, TDRL learns a zero cue value following extinction, thereby predicting that the dopaminergic cue response will reduce to zero concomitant with behavioral learning. However, ANCCR includes the measurement of a retrospective association between the cue and reward. This association does not update without rewards and hence, does not degrade due to extinction. This "long-term memory" was observed previously in orbitofrontal neurons projecting to the ventral tegmental area, the region where the somata of the mesolimbic dopamine neurons reside (*19*). Hence, the ANCCR hypothesis predicts that dopamine response will remain significantly positive long after animals learn to suppress anticipatory licking. This is because the cue remains a meaningful causal target despite extinction, even though the animals can learn extinction by noting that the base rate of rewards in the context becomes zero. Thus, Test 6 was whether dopamine cue response remained positive long after extinction was behaviorally learned (Fig. 4, J to L). As predicted by ANCCR but not RPE, dopamine cue response remained significantly positive well after animals cease to behaviorally respond to the cues (Fig. 4, J to L), consistent with prior studies (*48*, *49*).

To test whether the significant positive dopamine responses following extinction reflect a retrospective association between the cue and reward, we selectively reduced the retrospective association without reducing the prospective association. We maintained the fixed reward following the cue but added unpredictable rewards during the intertrial interval. In this experiment, not all rewards are preceded by the cue (i.e., retrospective association is weak), but all cues are followed by reward (i.e.,

prospective association is high). ANCCR predicts a rapid drop in dopamine cue response, but RPE predicts no change in cue response if TDRL only considers the cue-reward "trial period" (Test 7, fig. S10). The dopamine cue response remained significantly positive but decayed across trials faster than during extinction (Fig. 4, M to P).

Test 8 ("trial-less" cue-reward learning)

We performed another test related to the temporal scalability of TDRL versus retrospective causal inference (Test 8, Fig. 5). A key motivation for developing our model was that current TDRL models do not have a scalable representation of time, and hence fail to learn the correct structure of even simple environments in which a cue predicts a reward at a fixed delay with 100% probability (fig. S6). We devised an experiment in which a single cue predicted the reward at a fixed delay with 100% probability, but the cue occurred unpredictably with an exponentially distributed inter-cue interval between 0-99 s. We reduced the cue duration to 250 ms to allow nearby occurrences of the cue to be separated in time and had a long trace interval (3 s) following cue offset until reward delivery. Animals learned the cue-reward association quickly in this modified "trial-less" task (fig. S11).

In this task, a cue will occasionally be presented during the wait from the previous cue to its associated reward (Fig. 5A). If the "trial period" for cue-reward tasks is considered to be the interval between the cue and reward, the next "trial" can occasionally start before the previous trial is completed. During these "intermediate" cues, TDRL resets its prediction because it assumes a new trial has started without reward in the previous trial, thereby resulting in a negative RPE (i.e., the intermediate cue signals that the reward will now be further delayed; intuitively, the intermediate cue implies omission of reward after the previous cue). This results from the inability of TDRL to learn the correct structure of the task, which is that every cue occurrence causes a reward at a fixed delay (Supplementary Note 6).

On the other hand, ANCCR will learn that the intermediate cue is qualitatively similar to the previous cue because both predict reward, but due to a local increase in cue rate, ANCCR predicts a lower but positive response to the intermediate cue (Fig. 5, A and B). We did not observe any negative dopamine response to the intermediate cue regardless of how baseline was measured, and instead observed a positive but weaker response, consistent with ANCCR but not RPE (Fig. 5, C and D, and fig. S11).

Tests 9 to 11 (backpropagation within a trial)

A critical postulate of the TDRL RPE account is that dopamine responses drive value learning of the immediately preceding state. We tested three predictions of this central postulate that are each inconsistent with ANCCR. The first is that during the acquisition of trace conditioning, dopamine response systematically backpropagates from the moment immediately prior to reward

to the cue onset (50) (Test 9, Fig. 6A). Unlike TDRL RPE, ANCCR does not make such a prediction since delay periods are not broken into states in ANCCR. The second is that during sequential conditioning (cue1 predicts cue2 predicts reward), dopamine response first increases to cue2 and then increases to cue1 (Test 10, Fig. 6C). ANCCR instead predicts that dopamine responses to both cues will increase together and later diverge when cue2 is learned to be caused by cue1. The third is that artificially suppressing dopamine release from cue2 to reward during sequential conditioning will prevent learning of cue1 responses (Test 11, Fig. 6, E to H). In contrast, suppressing cue2 response in ANCCR only prevents the learning of the cue1→cue2 association and does not prevent the learning of cue1 response.

We tested the first prediction using the animals that underwent the previous cue-reward learning. Our observations were not consistent with a backpropagating bump of activity and were instead consistent with an increase in cue response over trials of learning (Fig. 6B) (see Supplementary Note 9 for potential reasons for discrepancy with a recent study). To test the second and third predictions, we performed sequential conditioning with an experimental group receiving inhibition of dopaminergic cell bodies from cue2 to reward, and a no-opsin control group that received the same laser but no inhibition of dopamine neurons. We measured NAcc dopamine release in both groups. The control group allowed us to test the dynamics of dopamine responses during sequential conditioning in the absence of dopamine neuron inhibition (i.e., the second prediction). Consistent with ANCCR, we experimentally found that cue2 and cue1 responses increased together early in learning prior to separating later in learning (Fig. 6D). To test the third prediction, we first verified robust inhibition of mesolimbic dopamine release during the cue2→reward delay in the experimental group (~0.6 times the reward response on day 1 of conditioning) (Supplementary Note 10). With such strong inhibition, TDRL RPE predicted no behavioral learning in this experiment, and a strong negative cue1 dopamine response (Fig. 6H and fig. S12). In contrast, ANCCR predicted largely intact learning of cue1, but with slower behavioral learning and reduced cue1 response (see Supplementary Note 10 for explanation). Consistent with ANCCR, we observed that every experimental animal learned the task and that mesolimbic dopamine acquired positive responses to cue1 in all experimental animals (Fig. 6I).

Discussion

The dynamics of mesolimbic dopamine release in NAcc were inconsistent with TDRL RPE across a multitude of experiments but remain consistent with a causal learning algorithm. The algorithm proposed here operates by testing whether a stimulus precedes reward beyond that expected by chance and by converting this association to a prospective prediction (Supplementary Note 7). Using this prediction, the algorithm learns a causal map of associations, and signals whether a stimulus has become a meaningful

causal target following such learning. Though our data are inconsistent with encoding of TDRL RPE by mesolimbic dopamine release, our framework is not inconsistent with prediction errors in general. Indeed, "prediction errors" related to event rates are a part of our framework (Supplementary Note 4).

The algorithm and results presented here provide a unified account of numerous published observations. Evidence across multiple species and brain regions shows that in addition to prospective associations, the brain stores memories of retrospective associations (*19*, *51*, *52*). Behavioral learning is also guided by retrospective associations (*18*, *53*). Dopamine responses remain significantly positive even to fully predicted, delayed rewards (*4*, *46*, *54–56*). This is usually explained by appealing to an internal uncertainty about the delay (*46*) but occurs without any accounting of temporal uncertainty in our theory (Fig. 2A). Consistent with our theory, a previous study observed no correlation between temporal uncertainty of an animal and the dopaminergic response to a fully predicted, delayed reward (*57*). Under some settings, dopamine reward responses during cue-reward conditioning have been observed to increase during initial learning, before decreasing back (*54*). While this observation is not consistent with RPE, it naturally results from our algorithm if the animal had no exposure to the reward in the experimental context prior to conditioning, as was the case (fig. S13). This might also explain why NAcc dopamine response to a predicted punishment might increase in some scenarios, while the responses to repeated punishments at fixed intervals decrease (*34*) (punishments are also meaningful causal targets; see Supplementary Note 8). ANCCR also explains recent observations of dopamine ramps used in favor of the RPE hypothesis (*58*) (fig. S13). Our explanation is also consistent with dopamine ramps in the striatum reflecting a causal association between an action and reward (*59*). Finally, dopamine responses guide learning in a way that sometimes violates the predictions of model-free TDRL (*17*, *60–63*). Our proposal that the dopaminergic system conveys whether cues are meaningful causal targets, thereby promoting the learning of their causes, explains these results (fig. S13).

Our work raises several questions for which reports in the literature suggests answers. First, how is retrospective cue-reward information conveyed to the dopaminergic system? Prior work suggests that the orbitofrontal cortex is a source of this information (*19*) (fig. S14). Second, how do animals infer the appropriate timescales in the world? Currently, we simply assume that animals set the appropriate timescale of an environment based on knowledge of the inter-reward interval. As a more principled solution, recent work has suggested that multiple parallel systems with different time constants exist in the brain and can learn a timescale invariant representation of past time (*64–67*). Third, are there as-yet unknown state space assumptions that make TDRL RPE fit our data? We cannot rule out all possible assumptions of TDRL state spaces because there is unlimited flexibility in

assuming the state space used by animals, thereby making them currently unfalsifiable (though see fig. S15). In the absence of such falsifiable assumptions, our work demonstrates that the TDRL algorithm with conventional state space assumptions does not explain the dynamics of dopamine release in NAcc. Fourth, does dopamine release in regions other than NAcc signal RPE? As mentioned in the introduction, we studied dopamine release in NAcc precisely because it is the region with the strongest support for the RPE hypothesis. Considering the theoretical advantages of ANCCR compared to TDRL RPE in learning associations between rates of events (figs. S6 and S15B), we believe that dopamine release in other regions might also be inconsistent with TDRL RPE; though, this remains to be tested. Finally, since it has been demonstrated that animal behavior and neural activity for even simple Pavlovian associations may be explained by the learning of causal cognitive maps (*68–71*), is all associative learning, including for action-conditional cognitive maps (*56*, *59*, *72–76*), the product of causal inference? This remains to be addressed. Collectively, our data demonstrate that mesolimbic dopaminergic signaling in NAcc is inconsistent with the dominant theory of TDRL RPE signaling and instead guides a causal learning algorithm.

REFERENCES AND NOTES
1. R. A. Rescorla, A. R. Wagner, "A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement" in *Classical conditioning II: Current Research and Theory*, pp. 64–99 (Appleton-Century-Crofts, 1972).
2. Y. Niv, G. Schoenbaum, Dialogues on prediction errors. *Trends Cogn. Sci.* **12**, 265–272 (2008). doi:10.1016/j.tics.2008.03.006 Medline
3. Y. Niv, Reinforcement learning in the brain. *J. Math. Psychol.* **53**, 139–154 (2009). doi:10.1016/j.jmp.2008.12.005
4. J. Y. Cohen, S. Haesler, L. Vong, B. B. Lowell, N. Uchida, Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature* **482**, 85–88 (2012). doi:10.1038/nature10754 Medline
5. W. Schultz, P. Dayan, P. R. Montague, A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997). doi:10.1126/science.275.5306.1593 Medline
6. Y. K. Takahashi, M. R. Roesch, R. C. Wilson, K. Toreson, P. O'Donnell, Y. Niv, G. Schoenbaum, Expectancy-related changes in firing of dopamine neurons depend on orbitofrontal cortex. *Nat. Neurosci.* **14**, 1590–1597 (2011). doi:10.1038/nn.2957 Medline
7. A. Mohebi, J. R. Pettibone, A. A. Hamid, J. T. Wong, L. T. Vinson, T. Patriarchi, L. Tian, R. T. Kennedy, J. D. Berke, Dissociable dopamine dynamics for learning and motivation. *Nature* **570**, 65–70 (2019). doi:10.1038/s41586-019-1235-y Medline
8. A. S. Hart, R. B. Rutledge, P. W. Glimcher, P. E. M. Phillips, Phasic dopamine release in the rat nucleus accumbens symmetrically encodes a reward prediction error term. *J. Neurosci.* **34**, 698–704 (2014). doi:10.1523/JNEUROSCI.2489-13.2014 Medline
9. J. J. Day, M. F. Roitman, R. M. Wightman, R. M. Carelli, Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens. *Nat. Neurosci.* **10**, 1020–1028 (2007). doi:10.1038/nn1923 Medline
10. P. E. M. Phillips, G. D. Stuber, M. L. A. V. Heien, R. M. Wightman, R. M. Carelli, Subsecond dopamine release promotes cocaine seeking. *Nature* **422**, 614–618 (2003). doi:10.1038/nature01476 Medline
11. M. P. Saddoris, F. Cacciapaglia, R. M. Wightman, R. M. Carelli, Differential Dopamine Release Dynamics in the Nucleus Accumbens Core and Shell Reveal Complementary Signals for Error Prediction and Incentive Motivation. *J. Neurosci.* **35**, 11572–11582 (2015). doi:10.1523/JNEUROSCI.2344-15.2015

Medline

12. M. P. Saddoris, J. A. Sugam, G. D. Stuber, I. B. Witten, K. Deisseroth, R. M. Carelli, Mesolimbic dopamine dynamically tracks, and is causally linked to, discrete aspects of value-based decision making. *Biol. Psychiatry* **77**, 903–911 (2015). doi:10.1016/j.biopsych.2014.10.024 Medline

13. H. M. Bayer, P. W. Glimcher, Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron* **47**, 129–141 (2005). doi:10.1016/j.neuron.2005.05.020 Medline

14. E. E. Steinberg, R. Keiflin, J. R. Boivin, I. B. Witten, K. Deisseroth, P. H. Janak, A causal link between prediction errors, dopamine neurons and learning. *Nat. Neurosci.* **16**, 966–973 (2013). doi:10.1038/nn.3413 Medline

15. C. Y. Chang, G. R. Esber, Y. Marrero-Garcia, H.-J. Yau, A. Bonci, G. Schoenbaum, Brief optogenetic inhibition of dopamine neurons mimics endogenous negative reward prediction errors. *Nat. Neurosci.* **19**, 111–116 (2016). doi:10.1038/nn.4191 Medline

16. H.-C. Tsai, F. Zhang, A. Adamantidis, G. D. Stuber, A. Bonci, L. de Lecea, K. Deisseroth, Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning. *Science* **324**, 1080–1084 (2009). doi:10.1126/science.1168878 Medline

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

18. C. R. Gallistel, A. R. Craig, T. A. Shahan, Contingency, contiguity, and causality in conditioning: Applying information theory and Weber's Law to the assignment of credit problem. *Psychol. Rev.* **126**, 761–773 (2019). doi:10.1037/rev0000163 Medline

19. V. M. K. Namboodiri, J. M. Otis, K. van Heeswijk, E. S. Voets, R. A. Alghorazi, J. Rodriguez-Romaguera, S. Mihalas, G. D. Stuber, Single-cell activity tracking reveals that orbitofrontal neurons acquire and maintain a long-term memory to guide behavioral adaptation. *Nat. Neurosci.* **22**, 1110–1121 (2019). doi:10.1038/s41593-019-0408-1 Medline

20. V. M. K Namboodiri, G. D. Stuber, The learning of prospective and retrospective cognitive maps within neural circuits. *Neuron* **109**, 3552–3575 (2021). doi:10.1016/j.neuron.2021.09.034 Medline

21. V. M. K. Namboodiri, How do real animals account for the passage of time during associative learning? *Behav. Neurosci.* **136**, 383–391 (2022). doi:10.1037/bne0000516 Medline

22. J. R. Platt, Strong Inference: Certain systematic methods of scientific thinking may produce much more rapid progress than others. *Science* **146**, 347–353 (1964). doi:10.1126/science.146.3642.347 Medline

23. G. Heymann, Y. S. Jo, K. L. Reichard, N. McFarland, C. Chavkin, R. D. Palmiter, M. E. Soden, L. S. Zweifel, Synergy of Distinct Dopamine Projection Populations in Behavioral Reinforcement. *Neuron* **105**, 909–920.e5 (2020). doi:10.1016/j.neuron.2019.11.024 Medline

24. W. Menegas, J. F. Bergan, S. K. Ogawa, Y. Isogai, K. Umadevi Venkataraju, P. Osten, N. Uchida, M. Watabe-Uchida, Dopamine neurons projecting to the posterior striatum form an anatomically distinct subclass. *eLife* **4**, e10032 (2015). doi:10.7554/eLife.10032 Medline

25. W. Menegas, K. Akiti, R. Amo, N. Uchida, M. Watabe-Uchida, Dopamine neurons projecting to the posterior striatum reinforce avoidance of threatening stimuli. *Nat. Neurosci.* **21**, 1421–1430 (2018). doi:10.1038/s41593-018-0222-1 Medline

26. S. Lammel, B. K. Lim, R. C. Malenka, Reward and aversion in a heterogeneous midbrain dopamine system. *Neuropharmacology* **76**, 351–359 (2014). doi:10.1016/j.neuropharm.2013.03.019 Medline

27. A. Lutas, H. Kucukdereli, O. Alturkistani, C. Carty, A. U. Sugden, K. Fernando, V. Diaz, V. Flores-Maldonado, M. L. Andermann, State-specific gating of salient cues by midbrain dopaminergic input to basal amygdala. *Nat. Neurosci.* **22**, 1820–1833 (2019). doi:10.1038/s41593-019-0506-0 Medline

28. B. T. Saunders, J. M. Richard, E. B. Margolis, P. H. Janak, Dopamine neurons create Pavlovian conditioned stimuli with circuit-defined motivational properties. *Nat. Neurosci.* **21**, 1072–1083 (2018). doi:10.1038/s41593-018-

29. A. A. Hamid, J. R. Pettibone, O. S. Mabrouk, V. L. Hetrick, R. Schmidt, C. M. Vander Weele, R. T. Kennedy, B. J. Aragona, J. D. Berke, Mesolimbic dopamine signals the value of work. *Nat. Neurosci.* **19**, 117–126 (2016). doi:10.1038/nn.4173 Medline

30. A. Lak, K. Nomoto, M. Keramati, M. Sakagami, A. Kepecs, Midbrain Dopamine Neurons Signal Belief in Choice Accuracy during a Perceptual Decision. *Curr. Biol.* **27**, 821–832 (2017). doi:10.1016/j.cub.2017.02.026 Medline

31. M. P. Saddoris, F. Cacciapaglia, R. M. Wightman, R. M. Carelli, Differential Dopamine Release Dynamics in the Nucleus Accumbens Core and Shell Reveal Complementary Signals for Error Prediction and Incentive Motivation. *J. Neurosci.* **35**, 11572–11582 (2015). doi:10.1523/JNEUROSCI.2344-15.2015 Medline

32. M. Darvas, A. M. Wunsch, J. T. Gibbs, R. D. Palmiter, Dopamine dependency for acquisition and performance of Pavlovian conditioned response. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 2764–2769 (2014). doi:10.1073/pnas.1400332111 Medline

33. K. Yamaguchi, Y. Maeda, T. Sawada, Y. Iino, M. Tajiri, R. Nakazato, S. Ishii, H. Kasai, S. Yagishita, A behavioural correlate of the synaptic eligibility trace in the nucleus accumbens. *Sci. Rep.* **12**, 1921 (2022). doi:10.1038/s41598-022-05637-6 Medline

34. M. G. Kutlu, J. E. Zachry, P. R. Melugin, S. A. Cajigas, M. F. Chevee, S. J. Kelly, B. Kutlu, L. Tian, C. A. Siciliano, E. S. Calipari, Dopamine release in the nucleus accumbens core signals perceived saliency. *Curr. Biol.* **31**, 4748–4761.e8 (2021). doi:10.1016/j.cub.2021.08.052 Medline

35. T. Patriarchi, J. R. Cho, K. Merten, M. W. Howe, A. Marley, W.-H. Xiong, R. W. Folk, G. J. Broussard, R. Liang, M. J. Jang, H. Zhong, D. Dombeck, M. von Zastrow, A. Nimmerjahn, V. Gradinaru, J. T. Williams, L. Tian, Ultrafast neuronal imaging of dopamine dynamics with designed genetically encoded sensors. *Science* **360**, eaat4422 (2018). doi:10.1126/science.aat4422 Medline

36. N. D. Daw, A. C. Courville, D. S. Touretzky, D. S. Touretzky, Representation and timing in theories of the dopamine system. *Neural Comput.* **18**, 1637–1677 (2006). doi:10.1162/neco.2006.18.7.1637 Medline

37. H.-E. Tan, A. C. Sisti, H. Jin, M. Vignovich, M. Villavicencio, K. S. Tsang, Y. Goffer, C. S. Zuker, The gut-brain axis mediates sugar preference. *Nature* **580**, 511–516 (2020). doi:10.1038/s41586-020-2199-7 Medline

38. C. R. Gallistel, T. A. Mark, A. P. King, P. E. Latham, The rat approximates an ideal detector of changes in rates of reward: Implications for the law of effect. *J. Exp. Psychol. Anim. Behav. Process.* **27**, 354–372 (2001). doi:10.1037/0097-7403.27.4.354 Medline

39. R. A. Rescorla, Pavlovian conditioning and its proper control procedures. *Psychol. Rev.* **74**, 71–80 (1967). doi:10.1037/h0024109 Medline

40. R. A. Rescorla, Probability of shock in the presence and absence of CS in fear conditioning. *J. Comp. Physiol. Psychol.* **66**, 1–5 (1968). doi:10.1037/h0025984 Medline

41. C. R. Gallistel, Robert Rescorla: Time, Information and Contingency. *Rev. Hist. Psicol.* **42**, 7–21 (2021).

42. V. M. K Namboodiri, T. Hobbs, I. Trujillo-Pisanty, R. C. Simon, M. M. Gray, G. D. Stuber, Relative salience signaling within a thalamo-orbitofrontal circuit governs learning rate. *Curr. Biol.* **31**, 5176–5191.e5 (2021). doi:10.1016/j.cub.2021.09.037 Medline

43. C. D. Fiorillo, W. T. Newsome, W. Schultz, The temporal precision of reward prediction in dopamine neurons. *Nat. Neurosci.* **11**, 966–973 (2008). doi:10.1038/nn.2159 Medline

44. A. Pastor-Bernier, A. Stasiak, W. Schultz, Reward-specific satiety affects subjective value signals in orbitofrontal cortex during multicomponent economic choice. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2022650118 (2021). doi:10.1073/pnas.2022650118 Medline

45. J. R. Hollerman, W. Schultz, Dopamine neurons report an error in the temporal prediction of reward during learning. *Nat. Neurosci.* **1**, 304–309 (1998). doi:10.1038/1124 Medline

46. S. Kobayashi, W. Schultz, Influence of reward delays on responses of dopamine neurons. *J. Neurosci.* **28**, 7837–7846 (2008). doi:10.1523/JNEUROSCI.1600-08.2008 Medline

47. M. E. Bouton, S. Maren, G. P. McNally, Behavioral and neurobiological mechanisms of pavlovian and instrumental extinction learning. *Physiol. Rev.* **101**, 611–681 (2021). doi:10.1152/physrev.00016.2020 Medline

48. W.-X. Pan, R. Schmidt, J. R. Wickens, B. I. Hyland, Tripartite mechanism of extinction suggested by dopamine neuron activity and temporal difference model. *J. Neurosci.* **28**, 9619–9631 (2008). doi:10.1523/JNEUROSCI.0255-08.2008 Medline

49. W. Zhong, Y. Li, Q. Feng, M. Luo, Learning and Stress Shape the Reward Response Patterns of Serotonin Neurons. *J. Neurosci.* **37**, 8863–8875 (2017). doi:10.1523/JNEUROSCI.1181-17.2017 Medline

50. R. Amo, S. Matias, A. Yamanaka, K. F. Tanaka, N. Uchida, M. Watabe-Uchida, A gradual temporal shift of dopamine responses mirrors the progression of temporal difference error in machine learning. *Nat. Neurosci.* **25**, 1082–1092 (2022). doi:10.1038/s41593-022-01109-2 Medline

51. K. E. Bouchard, M. S. Brainard, Neural encoding and integration of learned probabilistic sequences in avian sensory-motor circuitry. *J. Neurosci.* **33**, 17710–17723 (2013). doi:10.1523/JNEUROSCI.2181-13.2013 Medline

52. Y. Komura, R. Tamura, T. Uwano, H. Nishijo, K. Kaga, T. Ono, Retrospective and prospective coding for predicted reward in the sensory thalamus. *Nature* **412**, 546–549 (2001). doi:10.1038/35087595 Medline

53. H. E. Manzur, K. Vlasov, S.-C. Lin, A retrospective and stepwise learning strategy revealed by neuronal activity in the basal forebrain (2022), p. 2022.04.01.486795, (available at https://www.biorxiv.org/content/10.1101/2022.04.01.486795v1).

54. L. T. Coddington, J. T. Dudman, The timing of action determines reward prediction signals in identified midbrain dopamine neurons. *Nat. Neurosci.* **21**, 1563–1573 (2018). doi:10.1038/s41593-018-0245-7 Medline

55. K. Lee, L. D. Claar, A. Hachisuka, K. I. Bakhurin, J. Nguyen, J. M. Trott, J. L. Gill, S. C. Masmanidis, Temporally restricted dopaminergic control of reward-conditioned movements. *Nat. Neurosci.* **23**, 209–216 (2020). doi:10.1038/s41593-019-0567-0 Medline

56. B. Engelhard, J. Finkelstein, J. Cox, W. Fleming, H. J. Jang, S. Ornelas, S. A. Koay, S. Y. Thiberge, N. D. Daw, D. W. Tank, I. B. Witten, Specialized coding of sensory, motor and cognitive variables in VTA dopamine neurons. *Nature* **570**, 509–513 (2019). doi:10.1038/s41586-019-1261-9 Medline

57. R. N. Hughes, K. I. Bakhurin, E. A. Petter, G. D. R. Watson, N. Kim, A. D. Friedman, H. H. Yin, Ventral Tegmental Dopamine Neurons Control the Impulse Vector during Motivated Behavior. *Curr. Biol.* **30**, 2681–2694.e5 (2020). doi:10.1016/j.cub.2020.05.003 Medline

58. H. R. Kim, A. N. Malik, J. G. Mikhael, P. Bech, I. Tsutsui-Kimura, F. Sun, Y. Zhang, Y. Li, M. Watabe-Uchida, S. J. Gershman, N. Uchida, A Unified Framework for Dopamine Signals across Timescales. *Cell* **183**, 1600–1616.e25 (2020). doi:10.1016/j.cell.2020.11.013 Medline

59. A. A. Hamid, M. J. Frank, C. I. Moore, Wave-like dopamine dynamics as a mechanism for spatiotemporal credit assignment. *Cell* **184**, 2733–2749.e16 (2021). doi:10.1016/j.cell.2021.03.046 Medline

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

61. M. J. Sharpe, H. M. Batchelor, L. E. Mueller, C. Yun Chang, E. J. P. Maes, Y. Niv, G. Schoenbaum, Dopamine transients do not act as model-free prediction errors during associative learning. *Nat. Commun.* **11**, 106 (2020). doi:10.1038/s41467-019-13953-1 Medline

62. B. M. Seitz, I. B. Hoang, L. E. DiFazio, A. P. Blaisdell, M. J. Sharpe, Dopamine errors drive excitatory and inhibitory components of backward conditioning in an outcome-specific manner. *Curr. Biol.* **32**, 3210–3218.e3 (2022). doi:10.1016/j.cub.2022.06.035 Medline

63. I. Trujillo-Pisanty, K. Conover, P. Solis, D. Palacios, P. Shizgal, Dopamine neurons do not constitute an obligatory stage in the final common path for the evaluation and pursuit of brain stimulation reward. *PLOS ONE* **15**, e0226722 (2020). doi:10.1371/journal.pone.0226722 Medline

64. W. Z. Goh, V. Ursekar, M. W. Howard, Predicting the Future With a Scale-Invariant Temporal Memory for the Past. *Neural Comput.* **34**, 642–685 (2022). doi:10.1162/neco_a_01475 Medline

65. K. H. Shankar, M. W. Howard, A scale-invariant internal representation of time. *Neural Comput.* **24**, 134–193 (2012). doi:10.1162/NECO_a_00212 Medline

66. A. Tsao, J. Sugar, L. Lu, C. Wang, J. J. Knierim, M.-B. Moser, E. I. Moser, Integrating time from experience in the lateral entorhinal cortex. *Nature* **561**, 57–62 (2018). doi:10.1038/s41586-018-0459-6 Medline

67. W. Wei, A. Mohebi, J. D. Berke, Striatal dopamine pulses follow a temporal discounting spectrum (2021), p. 2021.10.31.466705, (available at https://www.biorxiv.org/content/10.1101/2021.10.31.466705v2).

68. T. J. Madarasz, L. Diaz-Mataix, O. Akhand, E. A. Ycu, J. E. LeDoux, J. P. Johansen, Evaluation of ambiguous associations in the amygdala by learning the structure of the environment. *Nat. Neurosci.* **19**, 965–972 (2016). doi:10.1038/nn.4308 Medline

69. S. J. Gershman, Y. Niv, Exploring a latent cause theory of classical conditioning. *Learn. Behav.* **40**, 255–268 (2012). doi:10.3758/s13420-012-0080-8 Medline

70. P. D. Balsam, C. R. Gallistel, Temporal maps and informativeness in associative learning. *Trends Neurosci.* **32**, 73–78 (2009). doi:10.1016/j.tins.2008.10.004 Medline

71. S. J. Gershman, D. M. Blei, Y. Niv, Context, learning, and extinction. *Psychol. Rev.* **117**, 197–209 (2010). doi:10.1037/a0017808 Medline

72. E. C. J. Syed, L. L. Grima, P. J. Magill, R. Bogacz, P. Brown, M. E. Walton, Action initiation shapes mesolimbic dopamine encoding of future rewards. *Nat. Neurosci.* **19**, 34–36 (2016). doi:10.1038/nn.4187 Medline

73. A. L. Collins, V. Y. Greenfield, J. K. Bye, K. E. Linker, A. S. Wang, K. M. Wassum, Dynamic mesolimbic dopamine signaling during action sequence learning and expectation violation. *Sci. Rep.* **6**, 20231 (2016). doi:10.1038/srep20231 Medline

74. A. Guru, C. Seo, R. J. Post, D. S. Kullakanda, J. A. Schaffer, M. R. Warden, Ramping activity in midbrain dopamine neurons signifies the use of a cognitive map (2020), p. 2020.05.21.108886, (available at https://www.biorxiv.org/content/10.1101/2020.05.21.108886v1).

75. N. G. Hollon, E. W. Williams, C. D. Howard, H. Li, T. I. Traut, X. Jin, Nigrostriatal dopamine signals sequence-specific action-outcome prediction errors. *Curr. Biol.* **31**, 5350–5363.e5 (2021). doi:10.1016/j.cub.2021.09.040 Medline

76. W. van Elzelingen, P. Warnaar, J. Matos, W. Bastet, R. Jonkman, D. Smulders, J. Goedhoop, D. Denys, T. Arbab, I. Willuhn, Striatal dopamine signals are region specific and temporally stable across action-sequence habit formation. *Curr. Biol.* **32**, 1163–1174.e6 (2022). doi:10.1016/j.cub.2021.12.027 Medline

77. J. P. Gavornik, M. G. H. Shuler, Y. Loewenstein, M. F. Bear, H. Z. Shouval, Learning reward timing in cortex through reward dependent expression of synaptic plasticity. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 6826–6831 (2009). doi:10.1073/pnas.0901835106 Medline

78. V. M. K. Namboodiri, S. Mihalas, T. M. Marton, M. G. Hussain Shuler, A general theory of intertemporal decision-making and the perception of time. *Front. Behav. Neurosci.* **8**, 61 (2014). doi:10.3389/fnbeh.2014.00061 Medline

79. P. N. Tobler, C. D. Fiorillo, W. Schultz, Adaptive coding of reward value by dopamine neurons. *Science* **307**, 1642–1645 (2005). doi:10.1126/science.1105370 Medline

80. N. F. Parker, C. M. Cameron, J. P. Taliaferro, J. Lee, J. Y. Choi, T. J. Davidson, N. D. Daw, I. B. Witten, Reward and choice encoding in terminals of midbrain dopamine neurons depends on striatal target. *Nat. Neurosci.* **19**, 845–854 (2016). doi:10.1038/nn.4287 Medline

81. H. Jeong, A. Taylor, J. R. Floeder, M. Lohmann, S. Mihalas, B. Wu, M. Zhou, D. A. Burke, V. M. K Namboodiri, Mesolimbic dopamine release conveys causal associations, Version 1, Zenodo (2022); https://zenodo.org/record/7302777#.Y4j503bMI2w

82. J. Pearl, Causal Diagrams for Empirical Research. *Biometrika* **82**, 669–688 (1995). doi:10.1093/biomet/82.4.669

83. C. R. Gallistel, J. Gibbon, Time, rate, and conditioning. *Psychol. Rev.* **107**, 289–344 (2000). doi:10.1037/0033-295X.107.2.289 Medline

84. I. M. Bright, M. L. R. Meister, N. A. Cruzado, Z. Tiganj, E. A. Buffalo, M. W. Howard, A temporal record of the past with a spectrum of time constants in the monkey entorhinal cortex. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 20274–20283 (2020). doi:10.1073/pnas.1917197117 Medline

85. Y. M. Ulrich-Lai, A. M. Christiansen, M. M. Ostrander, A. A. Jones, K. R. Jones, D. C. Choi, E. G. Krause, N. K. Evanson, A. R. Furay, J. F. Davis, M. B. Solomon, A. D. de Kloet, K. L. Tamashiro, R. R. Sakai, R. J. Seeley, S. C. Woods, J. P. Herman, Pleasurable behaviors reduce stress via brain reward pathways. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 20529–20534 (2010). doi:10.1073/pnas.1007740107 Medline

86. E. N. Holly, K. A. Miczek, Ventral tegmental area dopamine revisited: Effects of acute and repeated stress. *Psychopharmacology* **233**, 163–186 (2016). doi:10.1007/s00213-015-4151-3 Medline

87. C. E. Stelly, S. C. Tritley, Y. Rafati, M. J. Wanat, Acute Stress Enhances Associative Learning via Dopamine Signaling in the Ventral Lateral Striatum. *J. Neurosci.* **40**, 4391–4400 (2020). doi:10.1523/JNEUROSCI.3003-19.2020 Medline

88. D. George, R. V. Rikhye, N. Gothoskar, J. S. Guntupalli, A. Dedieu, M. Lázaro-Gredilla, Clone-structured graph representations enable flexible learning and vicarious evaluation of cognitive maps. *Nat. Commun.* **12**, 2392 (2021). doi:10.1038/s41467-021-22559-5 Medline

89. J. C. R. Whittington, T. H. Muller, S. Mark, G. Chen, C. Barry, N. Burgess, T. E. J. Behrens, The Tolman-Eichenbaum Machine: Unifying Space and Relational Memory through Generalization in the Hippocampal Formation. *Cell* **183**, 1249–1263.e23 (2020). doi:10.1016/j.cell.2020.10.024 Medline

90. B. Hayden, Y. Niv, The case against economic values in the brain (2020), doi:10.31234/osf.io/7hgup.

91. F. Etscorn, R. Stephens, Establishment of conditioned taste aversions with a 24-hour CS-US interval. *Physiol. Psychol.* **1**, 251–253 (1973). doi:10.3758/BF03326916

92. T. Akam, M. E. Walton, pyPhotometry: Open source Python based hardware and software for fiber photometry data acquisition. *Sci. Rep.* **9**, 3521 (2019). doi:10.1038/s41598-019-39724-y Medline

93. T. N. Lerner, C. Shilyansky, T. J. Davidson, K. E. Evans, K. T. Beier, K. A. Zalocusky, A. K. Crow, R. C. Malenka, L. Luo, R. Tomer, K. Deisseroth, Intact-Brain Analyses Reveal Distinct Information Carried by SNc Dopamine Subcircuits. *Cell* **162**, 635–647 (2015). doi:10.1016/j.cell.2015.07.014 Medline

94. A. Barr, M. Overduin, C. D. Proulx, Multi-Fiber Photometry to Record Neural Activity in Freely-Moving Animals, *J. Vis. Exp.* **152**, e60278 (2019). doi:10.3791/60278 Medline

95. E. A. Ludvig, R. S. Sutton, E. J. Kehoe, Stimulus representation and the timing of reward-prediction errors in models of the dopamine system. *Neural Comput.* **20**, 3034–3054 (2008). doi:10.1162/neco.2008.11-07-654 Medline

96. J. Pearl, Causal inference in statistics: An overview. *Stat. Surv.* **3**, 96–146 (2009). doi:10.1214/09-SS057

SUPPLEMENTARY MATERIALS

https://science.org/doi/10.1126/science.abq6740
Materials and Methods
Supplementary Notes 1 to 10
Figs. S1 to S15
Table S1
References (*82–96*)
MDAR Reproducibility Checklist

Date Accepted
Published online 8 Dec 2022
10.1126/science.abq6740
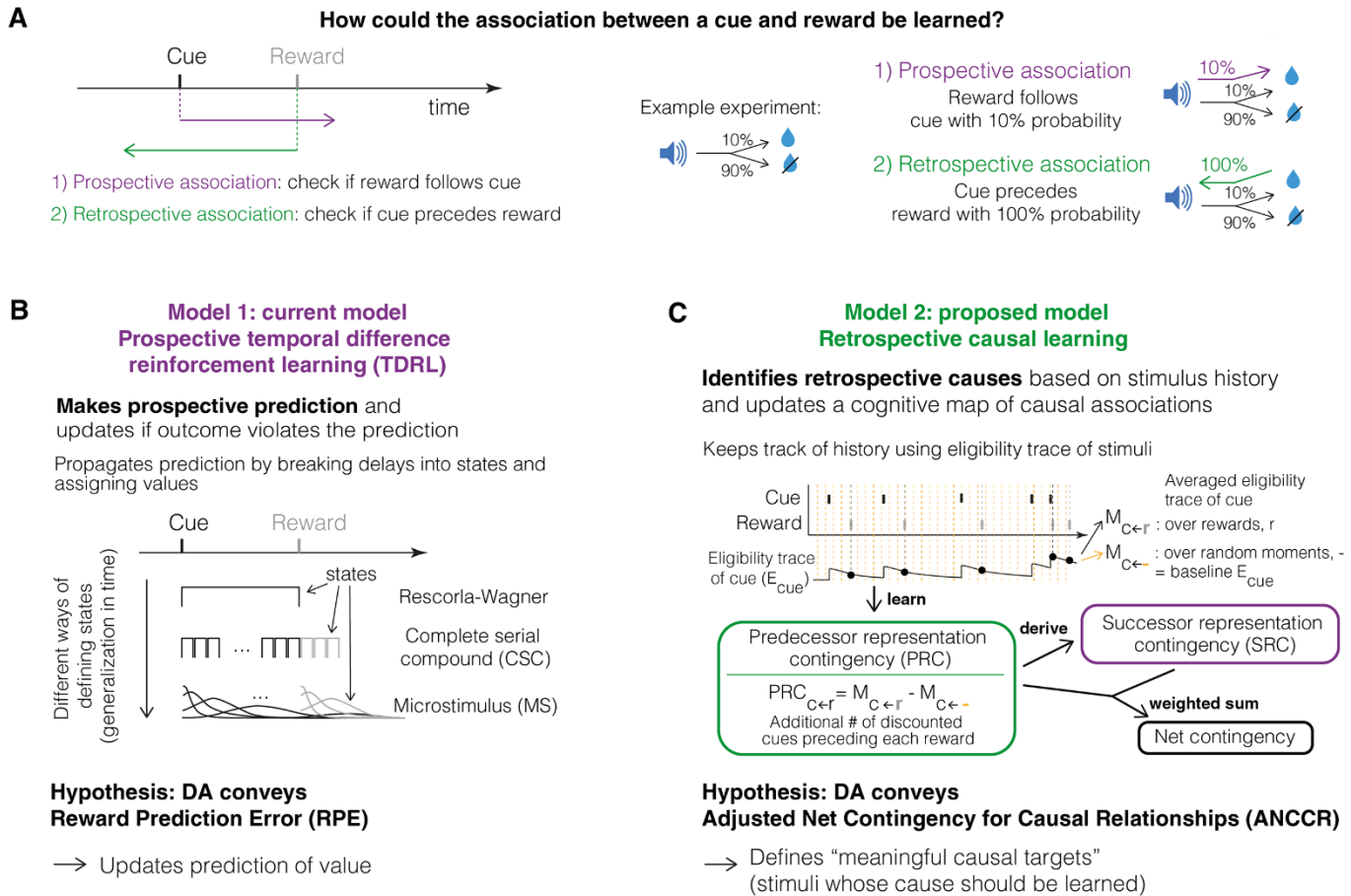
Fig. 1. An algorithm for uncovering causal associations in an environment. (**A**) Animals can learn cue-reward associations either prospectively ("does reward follow cue?") or retrospectively ("does cue precede reward?"). (**B**) The dominant model for cue-reward learning is temporal difference reinforcement learning, which learns the prospective association between a cue and reward, i.e., a measure of how often the reward follows the cue (cue value). To this end, the algorithm looks forward from a cue to predict upcoming rewards. When this prediction is incorrect, the original prediction is updated using a reward prediction error (RPE). The simplest of this family of models is the Rescorla-Wagner model which does not consider the delay between cue and reward. Temporal difference reinforcement learning (TDRL) algorithms extend this simple model to account for the cue-reward delay by modeling it as a series of states that measure time elapsed since stimulus onset. Two such examples are shown. (**C**) Here, we propose an algorithm which retrospectively learns the causes of meaningful stimuli such as rewards (figs. S1 to S4). Because causes precede outcomes, causal learning only requires a memory trace of the past. In our mechanistic model, a memory trace of prior stimuli is maintained using an exponentially-decaying eligibility trace for a stimulus (*77*), which allows the online calculation of the experienced rate of this stimulus (*78*). We hypothesized that mesolimbic dopamine activity signals ANCCR, a quantity that allows measuring whether an experienced stimulus is a meaningful causal target.
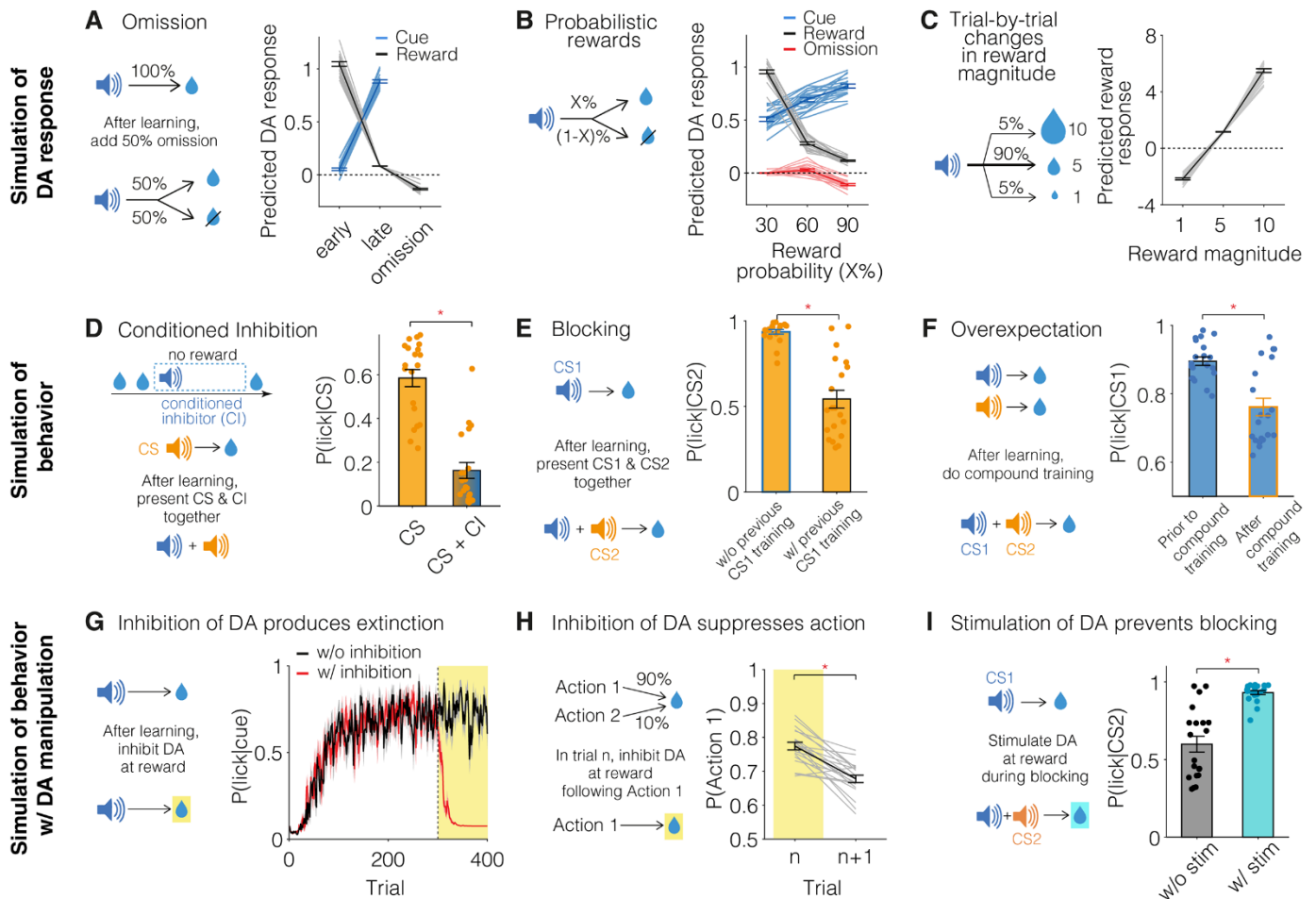
Fig. 2. The retrospective causal algorithm produces a signal similar to temporal difference reward prediction error (RPE) in simulations of previous experiments. (**A**) During simple conditioning of a cue-reward association, ANCCR appears qualitatively similar to an RPE signal, being low before and high after learning for the cue, whereas being high before and low after learning for the reward, and negative after omission of an expected reward. All error bars are standard error of the mean throughout the manuscript. (**B**) For probabilistic rewards, ANCCR produces qualitatively similar responses as RPE for cue, reward, and omission. Note that in (B) animals were never trained on a fully predicted reward. Slight differences in omission responses from [(A) to (B)] result from this difference. (**C**) For trial-by-trial changes in reward magnitude, ANCCR produces reward responses similar to positive and negative RPEs [similar to (*79*)]. (**D** to **F**) Simulations of ANCCR learning produces behavior consistent with conditioned inhibition (D), blocking (E), and overexpectation (F). (**G**) Simulated inhibition of dopamine at reward time in cue-reward conditioning produces extinction of learned behavior [(similar to (*55*)]. (**H**) Simulation of dopamine inhibition at reward time produces trial-by-trial changes in behavior [(similar to (*80*)]. (**I**) Simulation of unblocking due to dopamine activation at reward during blocking (similar to (*14*)).
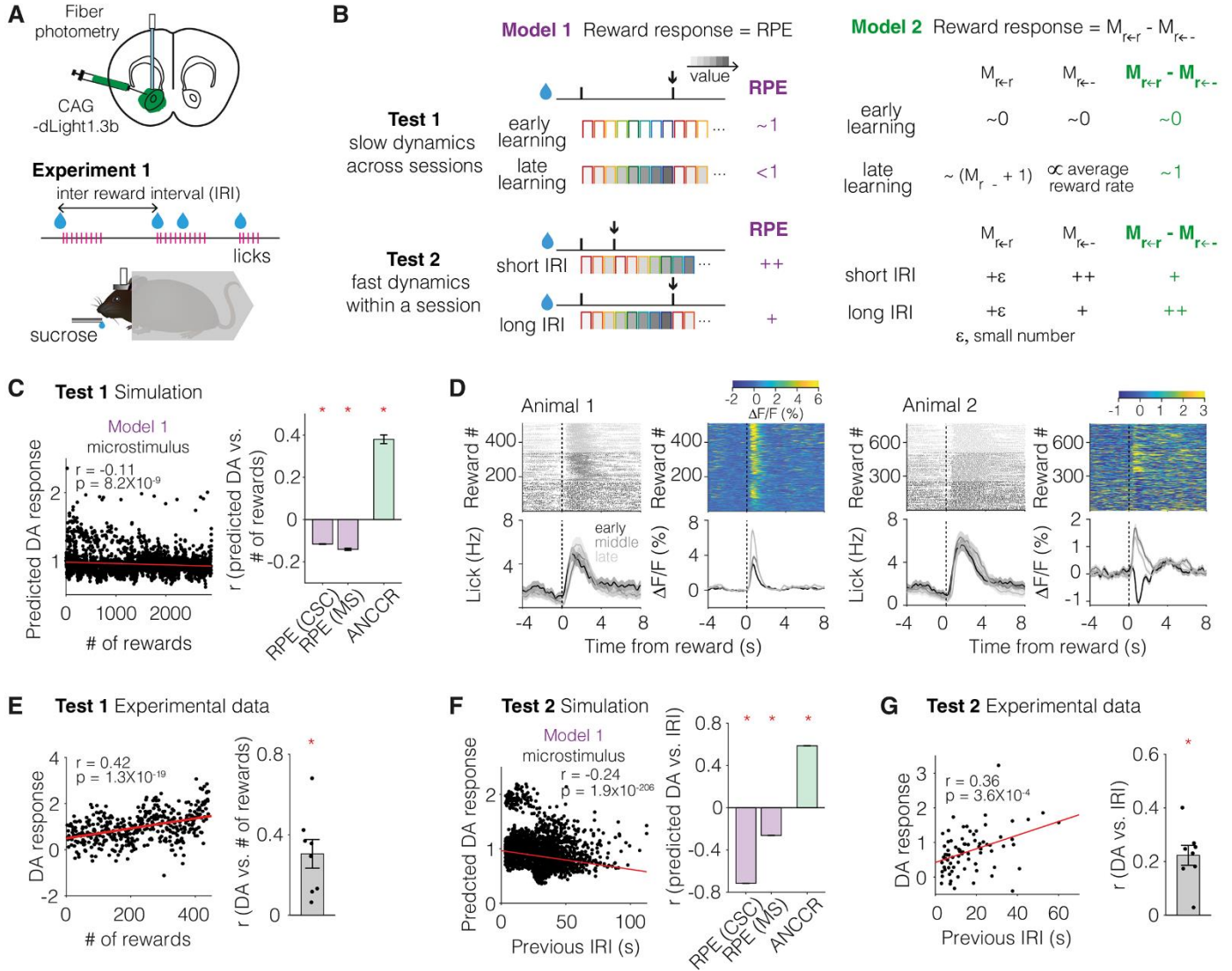
Fig. 3. The dynamics of dopamine responses to unpredicted rewards are consistent with ANCCR, but not TDRL RPE. (**A**) For the first two tests, we gave experimentally naïve mice random unpredictable sucrose rewards immediately following head-fixation while recording sub-second dopamine release in NAcc using the optical dopamine sensor, dLight 1.3b (Methods). Animals underwent multiple sessions with 100 rewards each (n = 8 mice). (**B**) Theoretical predictions for both models. Test 1: As a naïve animal receives unpredicted rewards, the RPE model predicts high responses since the rewards are unpredicted. Nevertheless, since the inter-reward interval (IRI) states acquire value over repeated experience, the RPE at reward will reduce with repeated experience. On the other hand, ANCCR predicts low reward responses early since an experimentally naïve animal will have no prior expectation/eligibility trace of sucrose early in the task but will subsequently approach a signal that is ~1 times the incentive value of sucrose. Test 2: The reward response following a short IRI will be larger in the RPE model because the reward was received earlier than expected, thereby resulting in a negative correlation between dopamine reward response and the previous IRI. However, since ANCCR has a subtractive term proportional to the baseline reward rate (M_r←in the figure), and baseline reward rate reduces with longer IRI, ANCCR predicts a positive correlation between dopamine reward response and the previous IRI. (**C**) Simulations confirming the intuitive reasoning from (B) for Test 1. CSC and MS stand for complete serial compound and microstimulus, respectively. (one sample t test against a null of zero; t(99) = RPE (CSC), −65.74; RPE (MS), −27.57; ANCCR, 18.60; Two-tailed P values = RPE (CSC), 1.7×10^{-83}, RPE (MS), 3.0×10^{-48}, ANCCR, 4.5×10^{-34}; n = 100 simulations). (**D**) Licking and dopamine response from two example mice (rewards with less than 3 s previous IRI were excluded to avoid confounding by ongoing licking responses). Though not our initial prediction, ANCCR can even account for the negative unpredicted sucrose response from Animal 2 (fig. S8). (**E**) Quantification of correlation between dopamine response and number of rewards. Left panel shows the data from an example animal and the right panel shows the population summary across all animals (one sample t test against a null of zero; t(7) = 4.40, two-tailed P = 0.0031; n = 8 animals). Reward response was defined as the difference of area under curve (AUC) of fluorescence trace between reward and baseline period (Methods). (**F**) Simulations confirming the intuitive reasoning from (B) for Test 2 (one sample t test against a null of zero; t(99) = RPE (CSC), -1.7×10^3, RPE (MS), −151.28, ANCCR, 335.03; Two-tailed P values = RPE (CSC), 5.0×10^{-223}, RPE (MS), 6.3×10^{-119}, ANCCR, 4.8×10^{-153}, n = 100 iterations). (**G**) Quantification of correlation between dopamine response and the previous IRI for an example session (left) and the population of all animals (one sample t test against a null of zero; t(7) = 5.95, two-tailed P = 5.7×10^{-4}, n = 8 animals). The average correlation across all sessions for each animal is plotted in the bar graph.

A

Model 1

RPE
↓
update the value
of the previous state
↓
in the next visit,
behave based on
updated value

hypothesis:
behavior learning
directly reflects
DA learning

Model 2

learn causal structure
↓
learn temporal delay
b/w causal events
↓
behave based on
learned temporal delay

hypothesis:
no one-to-one
relationship b/w
behavior learning
and DA learning

B

Trace interval

CS+ → Licks
CS−

Time (s)

C Experiment 2: Test 3 initial learning

Day 1 Day 2 Day 3 **Day 4** Day 7 Day 8 Day 9 **Day 12** Day 13 Day 14

Trial
Lick (Hz) CS+ CS−
Trial
ΔF/F (%)
Time from CS (s)

D Schematic for analysis

Y (artificial data)
Trial

cumsum(Y) / sum(Y) = Normalized cumsum (Y)
decreasing over time
increasing over time
Trial / max(trial) = Normalized trial

E DA considerably leads behavior

Normalized cumsum (lick rate)
Normalized trial

Normalized cumsum (DA cue response)
Normalized trial

F

Abruptness of change
Lick DA

Change trial
Lick DA

G Experiment 3: Tests 4-5
cue duration change 2s → 8s

Trial
Lick (Hz)
Time from CS (s)

Trial
ΔF/F (%)
2s CS+
early 8s CS+
middle 8s CS+
late 8s CS+
Time from CS (s)

H behavior changes but DA does not change

Normalized cumsum (median lick time)
Normalized trial

Normalized cumsum (DA cue response)
Normalized trial

I

Abruptness of change
Lick DA

ΔAIC (constant − changing)
Lick DA

J Experiment 4: Test 6
extinction

Trial
Lick (Hz)
before extinction
extinction
Time from CS (s)

Trial
ΔF/F (%)
Time from CS (s)

K DA remains positive despite extinction of behavior

Normalized cumsum (lick rate)
Normalized trial

Normalized cumsum (DA cue response)
Normalized trial

L

Abruptness of change
Lick DA

Change trial
Lick DA

M Experiment 5: Test 7

● reward predicted by CS+
◇ background reward

CS+ Licks
CS−
Time (s)

N

Experiment 4   Experiment 5
Prospective association decrease   no change
Retrospective association no change decrease

O

Trial
Lick (Hz)
w/o background
w/ background
Time from CS (s)

Trial
ΔF/F (%)
Time from CS (s)

P

Normalized cumsum (DA response)
Experiment 4
Experiment 5
Normalized trial

www.sciencemag.org

Fig. 4. The dynamics of dopamine responses during cue-reward learning are consistent with ANCCR, but not TDRL RPE.
(**A**) TDRL predicts that dopaminergic and behavioral learning will be tightly linked during learning. However, the causal learning model proposes that there is no one-to-one relationship between behavioral and dopaminergic learning. (**B**) Schematic of a cue-reward learning task in which one auditory tone predicted reward (labeled CS+) and another had no predicted outcome (labeled CS−). (**C**) Licking and dopamine measurements from an example animal showing that the dopamine response to CS+ significantly precedes the emergence of anticipatory licking (Days 4 vs 12 respectively, shown by the arrows). (**D**) Schematic to show a cumulative sum (cumsum) plot of artificial time-series data. A time-series that increases over trials appears below the diagonal in the cumsum plot with an increasing slope over trials, and one that decreases over trials appears above the diagonal. Further, a sudden change in timeseries appears as a sudden change in slope in the cumsum plot. (**E** and **F**) Dopamine cue response considerably leads behavior across animals. Each line is one animal, with the blue line corresponding to the example from (C). Behavioral learning is much more abrupt than dopaminergic learning (paired *t* test for abruptness of change; t(6) = 9.06; two-tailed $P = 1.0 \times 10^{-4}$; paired t test for change trial; t(6) = −2.93; two-tailed $P = 0.0263$; $n = 7$ animals). (**G**) Anticipatory licking and dopamine release in an example animal after increasing the cue duration from 2 s to 8 s while maintaining a 1 s trace interval and a long ITI (~33 s). Trials are shown in chronological order from bottom to top. The three vertical dashed lines indicate cue onset, cue offset, and reward delivery [also in (**J** and **O**)]. (**H** and **I**) Behavior is learned abruptly by all animals, but the dopaminergic cue response shows little to no change. The dashed vertical line is the trial at which the experimental condition transitions (in [**H**, **K**, and **P**]). We tested for the lack of change by showing that the Akaike Information Criterion (AIC) is similar between a model assuming change and a model assuming no change. Paired t test for abruptness of change; t(6) = 22.92; two-tailed $P = 4.52 \times 10^{-7}$; one-sample t test for ΔAIC against a null of zero; t(6) = 7.49 for lick, −0.86 for dopamine; two-tailed $P = 2.9 \times 10^{-4}$ for lick, 0.4244 for dopamine ($n = 7$ animals). (**J**) The dopaminergic cue response of an example animal remains positive well after it learns extinction of the cue-reward association. (**K** to **L**) Across all animals, the dopaminergic cue response remains significantly positive despite abrupt behavioral learning of extinction (paired *t* test for abruptness of change; t(6) = 5.67; two-tailed $P = 0.0013$; paired *t* test for change trial; t(6) = −2.40; two-tailed $P = 0.0531$; $n = 7$ animals). (**M**) Experiment to reduce retrospective association while maintaining prospective association. (**N**) Two experiments that show specific reduction in either prospective or retrospective association. (**O**) Licking and dopamine release from an example animal. (**P**) Dopamine cue response reduces more rapidly during the background reward experiment in which the cue is followed consistently by a reward than during extinction in which there is no reward (paired *t* test; t(6) = −3.51; two-tailed $P = 0.0126$; $n = 7$ animals).
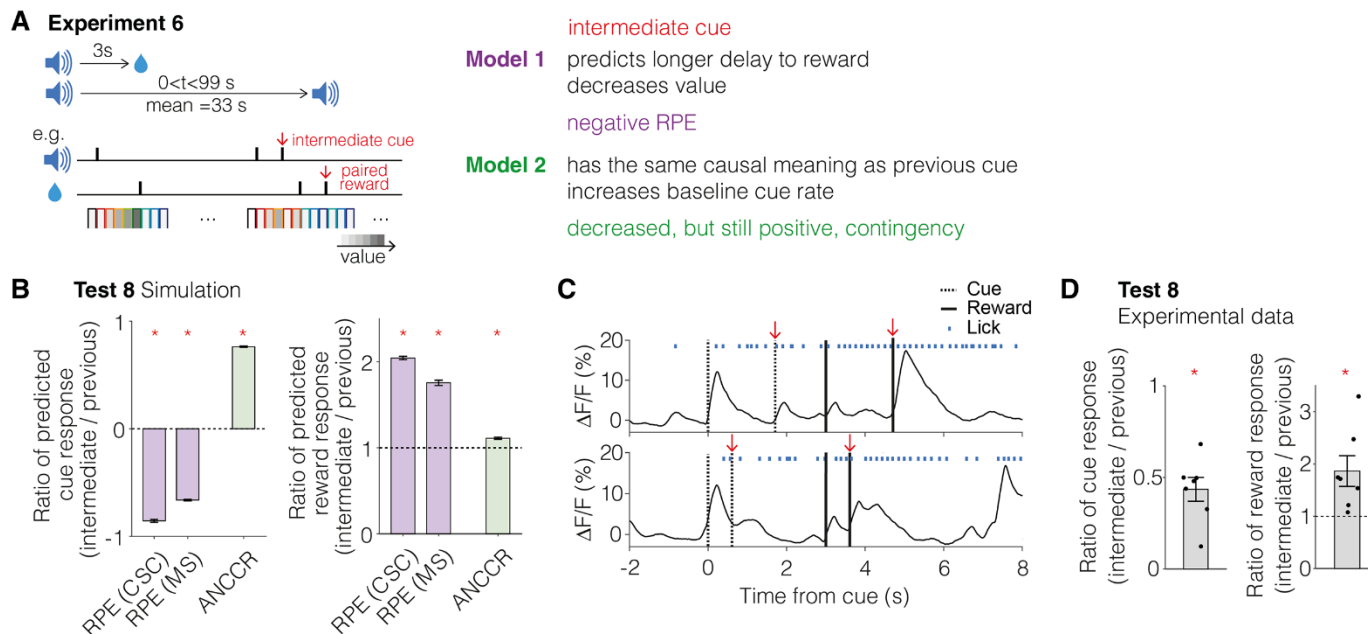
Fig. 5. Dopamine responses in a "trial-less" cue-reward task reflect causal structure like ANCCR, but unlike TDRL RPE. (**A**) A "trial-less" cue-reward learning task. Here, a cue (250 ms duration) is consistently followed by a reward at a fixed delay (3 s trace interval). However, the cues themselves occur with an exponential inter-cue interval with a 33 s mean. (**B**) Confirmation of these intuitions based on simulations (Methods) (One sample t test against a null of zero; t(99) = RPE (CSC), −114.74; RPE (MS), −181.32; ANCCR, 322.53; Two-tailed P values = RPE (CSC), 4.1×10^{-107}; RPE (MS), 1.1×10^{-126}; ANCCR, 2.1×10^{-151}; $n = 100$ iterations). Reward responses are predicted to be positive by both models (One sample t test against a null of one; t(99) = RPE (CSC), 87.67; RPE (MS), 62.86; ANCCR, 16.78; Two-tailed P values = RPE (CSC), 1.2×10^{-95}; RPE (MS), 1.3×10^{-81}; ANCCR, 1.1×10^{-30}; $n = 100$ iterations). (**C**) Example traces from one animal showing that the dopamine response to the intermediate cue is positive. (**D**) Quantification of the experimentally observed ratio between the intermediate cue response and the previous cue response (One sample t test against a null of zero; t(6) = 6.64, two-tailed P value = 5.6×10^{-4}; $n = 7$ animals), and reward response (One sample t test against a null of one; t(6) = 2.95; two-tailed P value = 0.0256; $n = 7$ animals).
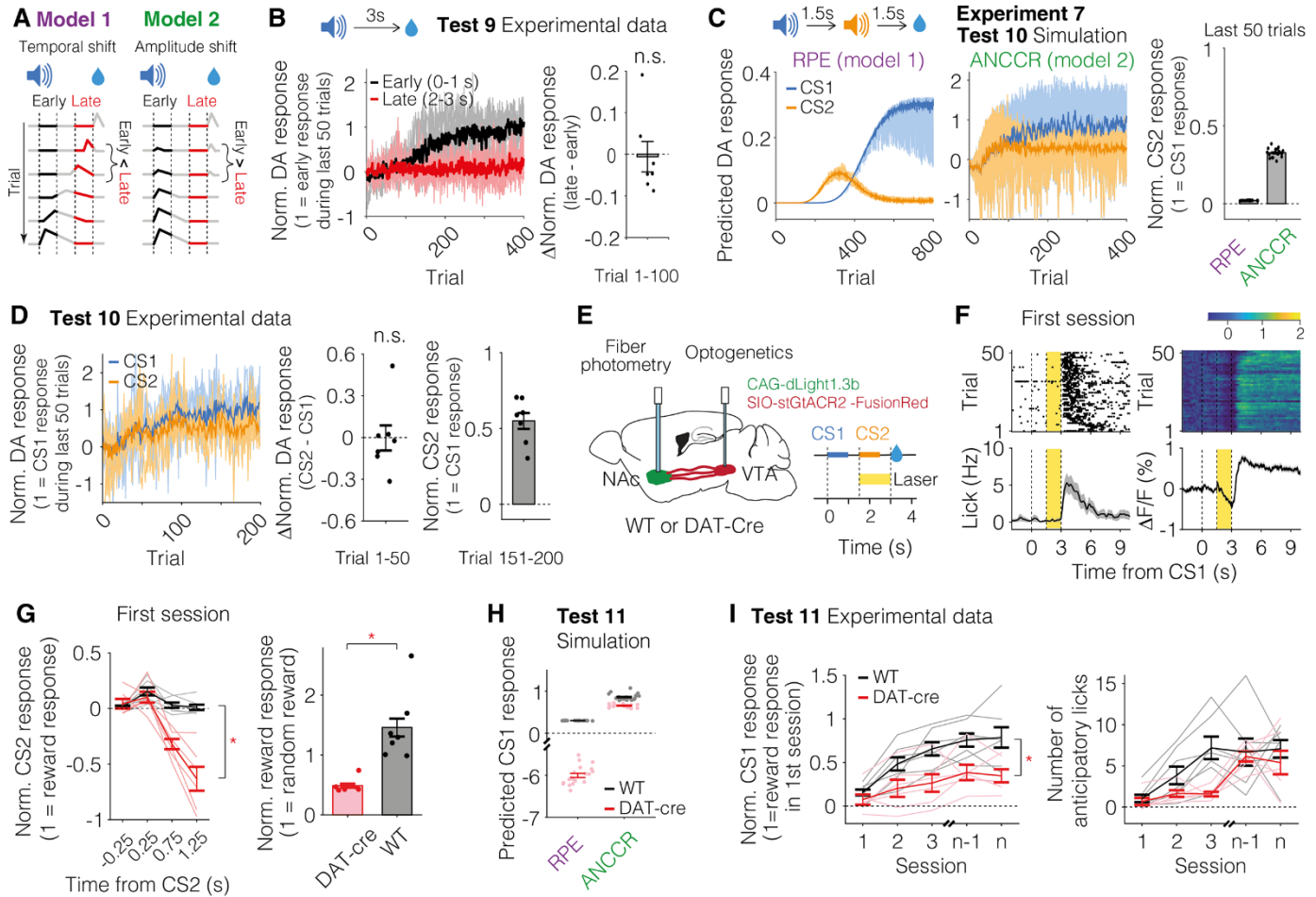
Fig. 6. No backpropagation of dopamine signals during learning. (**A**) Schematic of learning dynamics for pre-reward dopamine dynamics based on RPE or ANCCR signaling. Schematic was inspired from (*50*). If there is a temporal shift, the difference in dopamine response between early and late phases of a trial will be negative in the initial trials. (**B**) Dynamics of dopamine response during early and late periods within a trial over training (left), and their difference during first 100 trials. (**C**) Simulated dynamics for dopamine responses to cues (CS1 and CS2) during sequential conditioning (left), and averaged CS2 response during last 50 trials (right). (**D**) Experimental data showing dynamics of dopamine responses to cues (left). Response difference between two cues during early phase of learning (middle; similar to Fig. 6B, right) and CS2 response during late phase of learning (right, similar to Fig. 6C, right). (**E**) Schematic of optogenetic inhibition experiment during sequential conditioning for both experimental DAT-Cre animals receiving inhibition and control wild type animals receiving light but no inhibition. Animals received laser from CS2 until reward throughout conditioning. (**F**) Measured licking and dopamine responses on the first session of conditioning from an example experimental animal, showing robust inhibition. (**G**) Quantification of magnitude of inhibition during CS2 presentation prior to reward, and reward response. Both responses are measured relative to pre-CS1 baseline. (**H**) Predicted dopamine responses using simulations of RPE or ANCCR. (**I**) Experimental data showing CS1 response (left) and anticipatory licking (right) across sessions. Here, n represents the last session.

Science

Mesolimbic dopamine release conveys causal associations

Huijeong JeongAnnie TaylorJoseph R FloederMartin LohmannStefan MihalasBrenda WuMingkang ZhouDennis A BurkeVijay Mohan K Namboodiri

View the article online
https://www.science.org/doi/10.1126/science.abq6740
Permissions
https://www.science.org/help/reprints-and-permissions

Use of this article is subject to the Terms of service