NEUROSCIENCE

Mesostriatal dopamine is sensitive to changes in specific cue-reward contingencies

Eric Garr[1]*, Yifeng Cheng[1], Huijeong Jeong[2], Sara Brooke[3], Laia Castell[1], Aneesh Bal[1], Robin Magnard[1], Vijay Mohan K. Namboodiri[2,4], Patricia H. Janak[1,3,5]*

Learning causal relationships relies on understanding how often one event precedes another. To investigate how dopamine neuron activity and neurotransmitter release change when a retrospective relationship is degraded for a specific pair of events, we used outcome-selective Pavlovian contingency degradation in rats. Conditioned responding was attenuated for the cue-reward contingency that was degraded, as was dopamine neuron activity in the midbrain and dopamine release in the ventral striatum in response to the cue and subsequent reward. Contingency degradation also abolished the trial-by-trial history dependence of the dopamine responses at the time of trial outcome. This profile of changes in cue- and reward-evoked responding is not easily explained by a standard reinforcement learning model. An alternative model based on learning causal relationships was better able to capture dopamine responses during contingency degradation, as well as conditioned behavior following optogenetic manipulations of dopamine during noncontingent rewards. Our results suggest that mesostriatal dopamine encodes the contingencies between meaningful events during learning.

INTRODUCTION

Temporal contiguity between events—how close they occur in time—is not sufficient to explain learning. In appetitive Pavlovian conditioning in which animals learn to anticipate rewards based on antecedent cues, learning can be stunted by free rewards in the absence of the cue even when contiguity between cues and rewards is held constant (*1*–*3*). A similar phenomenon holds for aversive Pavlovian conditioning, in which animals fail to freeze in response to a cue that predicts shock when the shock is also delivered at the same rate in the absence of the cue (*4*). These findings have been used to argue that learning is a function of the cue-outcome contingency (*5*, *6*).

Despite the importance that contingency has for learning, investigations of midbrain dopamine (DA) function emphasize a role in learning that is based on contiguity. A prominent theory of midbrain DA function comes from temporal difference reinforcement learning (TDRL), which uses prediction errors to update state and action values (*7*). There is a remarkable resemblance between phasic DA activity and TDRL prediction errors (*8*–*11*), and exogenously evoked DA modulates behavior as if evoking an artificial prediction error (*12*–*15*). TDRL assumes that agents learn values of cues and actions, where value is defined as the time-discounted expectation of future reward (*7*). Put simply, the further away in time a reward is placed from some state, the less value will accrue to that state. Thus, value in TDRL is dependent on temporal contiguity between events, and many studies of how DA contributes to learning confound changes in value with changes in contingency. This is not to say that all variations of TDRL consider temporal contiguity as sufficient for learning (*16*), but rather as a necessary condition.

In the current study, we sought to hold the contiguity between cues and rewards constant while degrading the contingency for one pair of events but not another. This is known as outcome-selective contingency degradation. Two distinct cues were followed by distinct rewards, but one reward was also delivered noncontingently during the intertrial interval (ITI). From the point of view of a TDRL model that does not assign value to the ITI, the value of the degraded cue remains unchanged because it is followed by the same reward at the same interval and the same probability as the control cue. Even if the model assigns value to the ITI, under outcome-selective contingency degradation, the increased ITI value will affect both cues equally. TDRL thus predicts that outcome-selective contingency degradation should have no differential effects on degraded and nondegraded cues, both in terms of conditioned behavior and mesostriatal DA dynamics.

Here, we replicate the finding that outcome-selective contingency degradation gradually diminishes conditioned responding. Fiber photometry recordings of midbrain DA neuronal activity and of DA release in the ventral striatum reveal phasic responses to cues and rewards that diminish with contingency degradation, with the reward response losing its sensitivity to local reward history. Individual differences in performance are predicted by both the transient and prolonged components of DA cell body activity and neurotransmitter release during the cue-outcome interval. Optogenetic manipulations of DA neuron activity showed that noncontingent DA during the ITI is sufficient to attenuate conditioning responding, but unexpectedly, inhibition of a canonical mesostriatal DA projection does not interfere with learning contingency degradation. Last, we show that nearly all these results can be accounted for by a recently described computational model where DA guides retrospective causal learning (*17*).

RESULTS

Contingency degradation attenuates conditioned responding for reward

To manipulate cue-reward contingency while holding contiguity constant, we used a procedure known as Pavlovian contingency

[1]Department of Psychological & Brain Sciences, Krieger School of Arts & Sciences, Johns Hopkins University, Baltimore, MD 21218, USA. [2]Department of Neurology, University of California, San Francisco, CA 94158, USA. [3]Solomon H. Snyder Department of Neuroscience, Johns Hopkins School of Medicine, Baltimore, MD 21205, USA. [4]Weill Institute for Neurosciences, Kavli Institute for Fundamental Neuroscience, Center for Integrative Neuroscience, University of California, San Francisco, CA 94158, USA. [5]Kavli Neuroscience Discovery Institute, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA.
*Corresponding author. Email: egarr1@jh.edu (E.G.); patricia.janak@jhu.edu (P.H.J.)

degradation, wherein the contingency between a cue and a reward is diminished by delivering free rewards in the ITI. The design of the Pavlovian contingency degradation protocol was adapted from a previous report (*3*). Rats learned two cue-reward associations during Pavlovian acquisition. Two auditory cues (tone and white noise) were presented in separate trials at random times and in random order for 20 s followed by 0.5 probability of reward delivery (grain pellet or sucrose water). Each cue was associated with a different reward. During contingency degradation, cues were still followed by probabilistic rewards, but one reward was also delivered noncontingently during the ITI and at the same average rate as during the trial. This meant that the contingency between the reward delivered freely during the ITI and its associated cue was degraded, while the other cue-reward contingency remained intact. Put another way, one of the cues signaled a negligible change in reward rate [see description of cycle-to-trial (C/T) ratios below].

The conditioned response was defined as the anticipatory head entry into the food port during the 20-s cue before the trial outcome. During the acquisition phase of this behavioral procedure, the mean port entry rate increased during both cues equally, but during contingency degradation, responding became weaker during the degraded cue compared to the nondegraded cue (Fig. 1B). Repeated-measures analyses of variance (ANOVAs) on conditioned port entry rates with and without baseline entries subtracted revealed two-way interactions between cue type and training phase (all tests, $F_{1,34} > 18.24$, $P < 0.001$) and three-way interactions between cue type, session, and training phase (all tests, $F_{7,238} > 2.64$, $P < 0.013$). Post hoc contrasts revealed a difference between cue types during the last four combined sessions of contingency degradation ($P < 0.05$). There was also a main effect of the training phase whether baseline was subtracted or not (all tests, $F_{1,34} > 4.99$, $P < 0.033$), indicating that, in addition to cue-selective effects, contingency degradation also generally weakened conditioned responding. The effect of contingency degradation was not mediated by context, which rules out a value-based account of contingency degradation whereby the value of the context overshadows the value of the degraded cue (fig. S1A).

To confirm that contingency degradation affected learning and not just performance, a reacquisition session was run in which the original contingencies from the acquisition phase were reinstated—noncontingent rewards were no longer delivered during the ITI (Fig. 1B). The mean entry rate was significantly higher during the nondegraded cue whether baseline was subtracted ($t_{26} = 2.17$, $P = 0.04$) or not ($t_{26} = 2.58$, $P = 0.016$), indicating a lasting impact of the outcome-specific contingency degradation procedure.

Next, we examined how local trial history affected conditioned port entries. This was achieved by modeling the number of cued port entries as a linear combination of trial outcomes up to two trials back in time. The regression coefficients and intercepts were combined for each rat to yield the predicted number of cued entries as a function of trial history (Fig. 1C). A repeated-measures ANOVA revealed a main effect of reward history ($F_{1,34} = 4.23$, $P = 0.047$), a main effect of cue type ($F_{1,34} = 5.57$, $P = 0.024$), and a cue type × reward history interaction ($F_{1,34} = 4.11$, $P = 0.05$). Post hoc contrasts showed that two consecutively rewarded trials increased the number of port entries more for the subsequent nondegraded cue ($P < 0.05$). However, when a rewarded trial was preceded by an omission trial, the subsequent cue port entries did not significantly differ between cue types ($P > 0.05$). This analysis shows that behavioral sensitivity to local reward history was weaker for the cue that underwent contingency degradation.

To confirm that the nondegraded cue signaled a larger change in reward rate, the C/T ratio was computed (Fig. 1D). The "cycle" is the mean interval between all deliveries of one reward type, and the "trial" is the mean interval between that same reward and its preceding cue during a trial (*18*). During contingency degradation, the C/T ratio was significantly smaller for the degraded versus nondegraded cue ($t_{26} = 13.90$, $P < 0.001$). Collectively, these analyses show that conditioned head entry rates during a Pavlovian cue were selectively attenuated by outcome-selective contingency degradation.

Contingency degradation attenuates the VTA DA neuron response to cue and reward

To determine the impact of contingency degradation on the encoding of cue-reward contingencies, we measured DA neuron activity in the ventral tegmental area (VTA) when one cue-reward contingency was degraded while the other remained intact. Tyrosine hydroxylase (TH)–Cre rats ($n = 13$) were injected with a Cre-dependent GCaMP6f virus in the VTA for imaging of calcium fluorescence via fiber photometry. The $\Delta F/F$ signal was aligned to cue onset, reward delivery, and reward omission (reward is delivered after each cue with 0.5 probability) during the last session of contingency degradation (Fig. 2B; see fig. S2 for example traces). The DA response to the nondegraded cue was significantly higher than to the degraded cue ($t_{12} = 2.36$, $P = 0.037$), and the DA response to reward onset was greater following the nondegraded cue than the degraded cue ($t_{12} = 3.82$, $P = 0.002$). The DA response to reward omission was divided into a positive phase, when the signal rose above baseline, and a negative phase when the signal dropped below baseline (see Materials and Methods). The omission response did not differ in either the positive ($t_{12} = 2.13$, $P = 0.055$) or negative phase ($t_{12} = 0.84$, $P = 0.42$) when comparing the degraded and nondegraded outcome pair. During the first session of contingency degradation, the reward responses differed in the same direction but the cue responses did not, indicating that it is the cue-evoked component that changes with conditioned responding (fig. S3). These findings show that DA neuron responses to cues and rewards diminish during contingency degradation and that these changes are selective to the target cue-reward association.

We ran additional analyses to explore whether the observed DA signal conformed to a canonical reward prediction error signal. One way to test for reward prediction error encoding is to quantify how the neural response to trial outcome is affected by local reward history (*19*–*21*). This was achieved by modeling the photometry response to the trial outcome as a linear combination of trial outcomes starting from any given trial and going two trials back. The regression coefficients and intercepts were combined for each rat to yield the predicted photometry response as a function of trial history (Fig. 2C). A repeated-measures ANOVA revealed a main effect of reward history ($F_{2,24} = 40.83$, $P < 0.001$), no main effect of cue type, and a cue type × reward history interaction ($F_{2,24} = 15.48$, $P < 0.001$). Post hoc contrasts showed that the outcome response following the nondegraded cue resembled the pattern that would be expected if DA neurons were computing prediction errors. Specifically, the photometry response was greatest when the current trial was rewarded but the previous trial was not, followed by a smaller response when the reward was delivered two trials in a row, and an even smaller response when the current trial was not rewarded but the previous one was ($P < 0.05$). This was not true of the outcome response following

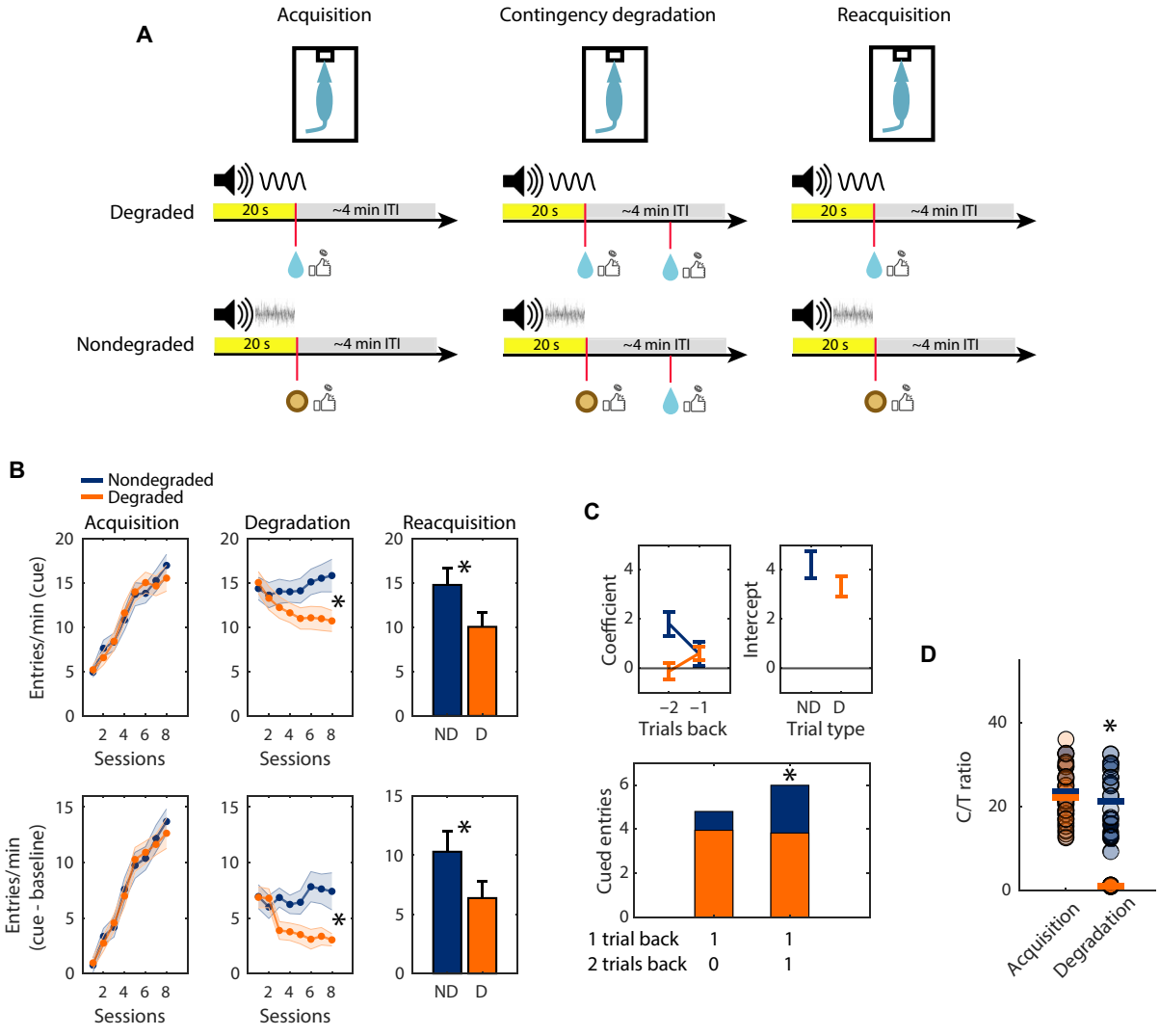

Fig. 1. Conditioned responding during Pavlovian contingency degradation. (**A**) Design of Pavlovian contingency degradation experiments showing the two trial types with different food rewards. Trials were presented in random order. The specific cue-outcome and noncontingent outcome assignments are shown for illustration purposes, but they were counterbalanced across rats. Coin flip means 0.5 probability of reward. (**B**) Mean (±SEM) port entry rates during the 20-s cue period (top) and with baseline subtracted (bottom) for each phase of the experiment. *$P < 0.05$. Data are pooled across 35 rats from three separate experiments (VTA GCaMP, NAc dLight, and context change/reward preference), except during reacquisition, which only includes rats from the VTA GCaMP and NAc dLight experiments. (**C**) Top: Mean (±SEM) regression coefficients (left) and intercepts (right) when the number of cued port entries was regressed against trial outcome history, shown separately for nondegraded (ND) and degraded (D) trials. Bottom: Mean predicted entry rates on trial n as a function of the outcome on trials $n - 1$ and $n - 2$. 1 indicates reward and 0 indicates omission. *$P < 0.05$. All data are taken from the last session of contingency degradation. (**D**) C/T ratios calculated from the final sessions of acquisition and contingency degradation. *$P < 0.05$. Circles are individual rats and horizontal bars are means. Data are pooled across 27 rats from the VTA GCaMP and NAc dLight experiments. VTA, ventral tegmental area; NAc, nucleus accumbens.

degraded trials ($P > 0.05$). The same analysis was applied to the photometry response to cue onset, with no main effects or interaction (fig. S5A).

A reward prediction error account of DA signaling predicts that the reward delivered during the ITI should elicit a larger response compared to when it is delivered after a fixed-duration cue. We compared the VTA DA response aligned to reward delivery at the end of a trial—specifically, at the offset of the degraded cue—and to delivery during the ITI (Fig. 2D). Notably, the reward identity is the same in both cases. The response to the reward delivered during the ITI was larger than to the same reward delivered at the end of a trial ($t_{12} = 3.56$, $P = 0.004$).

Contingency degradation attenuates NAc DA release to cue, reward delivery, and reward omission

To confirm that these findings hold for DA neurotransmitter release, wild-type rats ($n = 14$) were injected with a dLight1.2 virus in the core of the nucleus accumbens (NAc) for imaging of DA release via fiber photometry. The z-scored $\Delta F/F$ signal aligned to trial events is shown in Fig. 3B. Similar to VTA DA neuron activity, DA release aligned to the nondegraded cue was significantly higher than to the degraded cue ($t_{13} = 3.18$, $P = 0.007$), and DA release aligned to reward onset was greater following the nondegraded cue than the degraded cue ($t_{13} = 5.54$, $P < 0.001$). The reward omission response did not differ during the positive phase ($t_{13} = 1.67$, $P = 0.118$) but

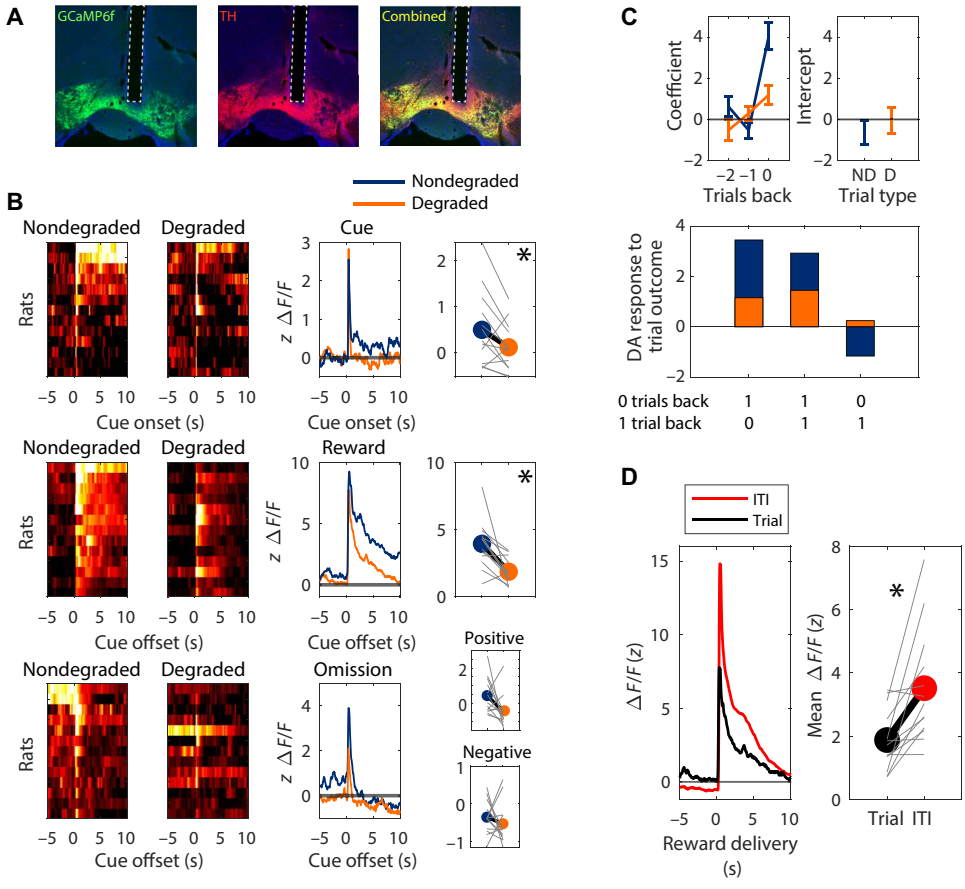

Fig. 2. Fiber photometry recordings in VTA DA neurons during contingency degradation. (**A**) Coronal section from one TH-Cre rat injected with DIO-GCaMP6f in the VTA. Dashed lines outline the fiber tract. (**B**) Left column: Heatmaps showing GCaMP response aligned to trial events separated by nondegraded and degraded trials. Each row shows the mean for each rat (averaged across trials). Middle column: Mean photometry traces, averaged across rats, in response to trial events, shown separately for nondegraded (blue) and degraded (orange) trials. Right column: Mean *z*-scored Δ*F*/*F* in response to trial events for nondegraded and degraded trials. **P* < 0.05. Reward omission means were quantified separately for the positive (top) and negative (bottom) phases of the signal. Gray lines show individual rats and large circles show means across rats. All data are from the last session of contingency degradation. (**C**) Top: Mean (±SEM) regression coefficients (left) and intercepts (right), averaged across rats, when the GCaMP response to the trial outcome was regressed against trial outcome history, shown separately for nondegraded and degraded trials. Bottom: Mean predicted GCaMP response at the time of the outcome on trial *n* as a function of the outcome on trial *n* and trial *n* − 1. 1 indicates reward and 0 indicates omission. All data are taken from the last session of contingency degradation. (**D**) Mean GCaMP signal, averaged across rats, aligned to the same reward delivered during the ITI or at the end of a trial at the offset of the degraded cue. **P* < 0.05. Gray lines represent individual rats. All data are from the last session of contingency degradation.

was significantly different during the negative phase (t_{13} = 2.45, P = 0.029), with a greater drop below baseline after termination of the nondegraded cue. During the first session of contingency degradation, the reward responses differed in the same direction but the cue responses did not (fig. S4).

As above, DA release during the trial outcome was modeled as a combination of current and past trial outcomes (Fig. 3C). The predicted response to trial outcomes as a function of trial history showed a main effect of reward history ($F_{2,26}$ = 15.31, P < 0.001), a main effect of cue type ($F_{1,13}$ = 9.48, P = 0.009), and a cue type × reward history interaction ($F_{2,26}$ = 7.55, P = 0.003). Post hoc contrasts showed that the outcome response following the nondegraded cue was greatest when the current trial was rewarded but the previous trial was not, followed by a smaller response when the reward was delivered two trials in a row, and an even smaller response when the current trial was not rewarded but the previous one was (P < 0.05). This was not true of the outcome response following degraded trials (P > 0.05). This analysis shows that the modulation of NAc DA by trial outcomes accords with TDRL only when the outcome is contingent on the cue, as observed above for the VTA GCaMP analysis.

We next compared DA release aligned to reward delivery at the end of a trial and to delivery during the ITI (Fig. 3D). The mean response to the reward delivered during the ITI did not differ from the response to the same reward delivered at the end of a trial (t_{13} = 0.34, P = 0.743), but the peak response did (t_{13} = 2.96, P = 0.011). Overall, the measurements of DA cell body activity and DA terminal release reveal a highly similar pattern of changes in cue- and reward-elicited activity following contingency degradation that are specific to the targeted cue-outcome pair. These findings confirm that DA signaling is sensitive to changes in contingency.

Neither behavior nor DA transients are explained by satiety

During contingency degradation, one reward type is delivered more frequently than the other, and it is possible that rats could become

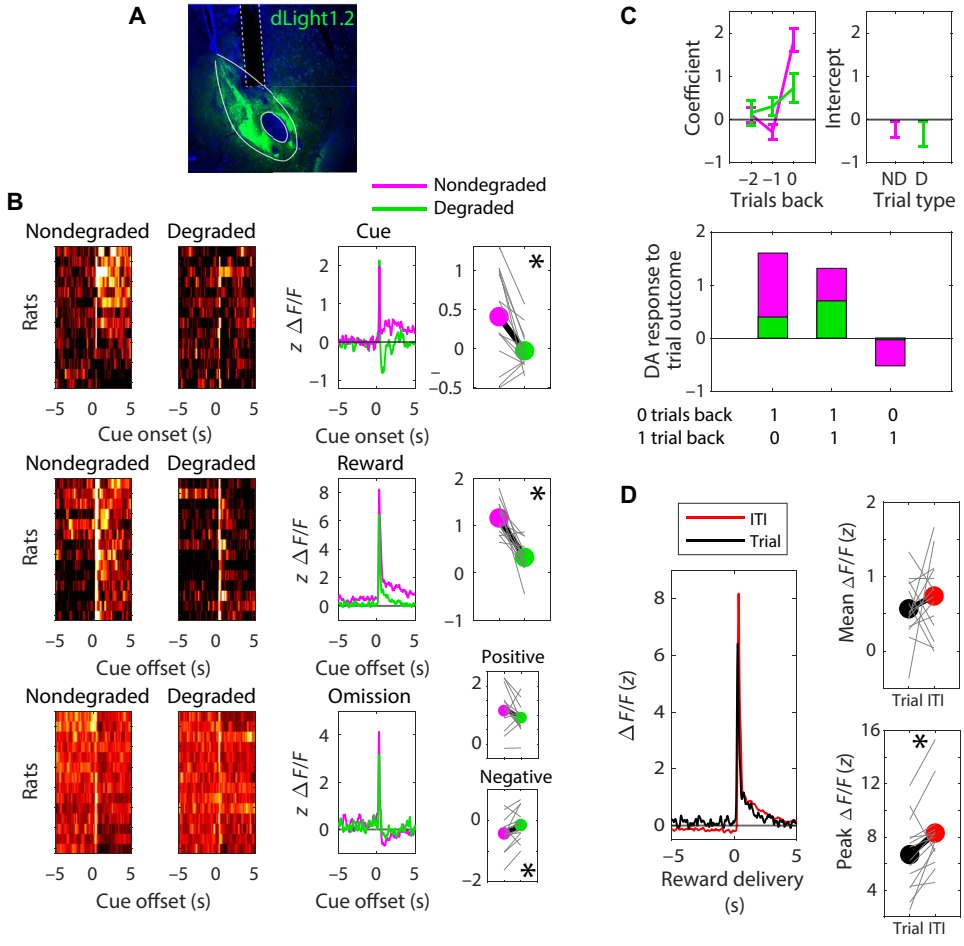

Fig. 3. Fiber photometry recordings of DA release in the NAc during contingency degradation. (**A**) Coronal section from one rat injected with dLight1.2 in the NAc core. Dashed line outlines the fiber tract. (**B**) Same as Fig. 2, except that nondegraded trials are in magenta and degraded trials are in green. (**C**) Same as in Fig. 2. (**D**) Same as in Fig. 2 but also showing the peak z Δ*F*/*F* signal (right). All data are from the last session of contingency degradation.

sated on that reward and this could differentially affect both the behavioral response and the DA response to trial events. However, reward preference tests showed that contingency degradation induced only general satiety, not specific satiety, indicating that satiety cannot account for the differences in behavioral and neural responses to events during degraded and nondegraded trials (fig. S1B). To test whether the photometry response to trial events was affected by satiety, we regressed the photometry response to cues, reward deliveries, and reward omissions on the times at which those events occurred during the last degradation session. The assumption is that satiety should grow over the course of the session. The times at which cues, reward deliveries, and reward omission occurred did not affect the photometry responses to those events for either VTA DA neuron or NAc DA release recordings (fig. S6; one-sampled *t* test $P > 0.05$). An additional analysis of the DA reward response as a function of trial (first and last) and cue type (nondegraded and degraded) revealed only a main effect of cue type ($P < 0.003$), but no main effect of trial number nor an interaction. These results show that DA signals do not progressively change over session time as might be expected if satiety were contributing to the observed DA responses.

Transient and prolonged DA during the cue-outcome interval predicts conditioned responding

We observed that degrading the cue-reward relationship by adding noncontingent reward deliveries altered DA responses to both cue and reward and altered behavioral responding to the degraded cue. To relate the neural dynamics to behavioral performance, we first regressed the rate, timing, and latency of conditioned entries against the photometry responses to cues and trial outcomes on a trial-by-trial basis. These analyses did not yield any meaningful patterns, suggesting that the magnitude of DA responses did not determine the performance features of the conditioned response on a trial-by-trial basis (figs. S7 and S8). We then asked if there was a reliable relationship between the relative behavioral responses to the two cues and the neural responses to the two cues at the level of the session rather than individual trials. Unlike the trial-by-trial analysis, which requires using only the transient component of the cue-evoked response to predict upcoming port entry features, an analysis that relates session-averaged behavior and neural responses allowed us to segment the cue-evoked responses into transient and prolonged components to predict session-level behavior (see Materials and Methods). Rats were split into two groups: those that showed a

behavioral effect of contingency degradation and those that did not. Groups were determined by subtracting the mean port entry rate during the degraded cue from the mean port entry rate during the nondegraded cue and using the 0 boundary to divide the groups. Given the small number of rats that did not show an effect of contingency degradation, analyses were combined across GCaMP and dLight experiments (see Fig. 4, A and B, for individual experiment data). The cue-evoked responses were segmented into transient and prolonged components. These components were both predictive of behavioral performance (see fig. S9 for analysis of reward and omission responses, which were not predictive of performance).

For the transient component of the DA response, a cue × group ANOVA yielded a significant interaction ($F_{1,25} = 10.53$, $P = 0.003$; Fig. 4C, top). Regressing the session-averaged difference in the entry rates between nondegraded and degraded trials against the difference between the transient signals revealed a significantly positive relationship ($\beta = 2.76$, $t_{25} = 2.61$, $P = 0.015$; Fig. 4C, bottom). For the prolonged component of the DA response, a cue × group ANOVA yielded an interaction that was shy of significance ($F_{1,25} = 3.52$, $P = 0.072$; Fig. 4D, top), but regressing the difference in the port entry rates between nondegraded and degraded trials against the difference between prolonged signals also revealed a significantly positive relationship ($\beta = 5.90$, $t_{25} = 2.33$, $P = 0.028$; Fig. 4D, bottom). Overall, these data indicate that the transient and prolonged DA signals during the cue-outcome interval are sensitive to cue-reward contingency and may drive individual differences in Pavlovian learning and motivation.

Inhibiting DA activity and release during ITI rewards does not block the effect of contingency degradation on conditioned responding

During fiber photometry recordings, noncontingent food rewards evoked a strong transient response in VTA DA neurons and DA release in the NAc. To test whether noncontingent DA activity is necessary for rats to learn about contingency degradation, TH-Cre rats ($n = 10$) underwent optogenetic inhibition of DA neurons during noncontingent reward deliveries. Green laser stimulation was targeted to halorhodopsin (eNpHr)-expressing DA neurons in the VTA only during noncontingent rewards during the ITI. A control group of TH–Cre-negative rats ($n = 9$) also received the same conditioning and laser treatments. Conditioned port entry rates during the cues across acquisition and degradation are shown in Fig. 5C. Rats in both groups showed lower conditioned responding during the degraded versus nondegraded cue (training phase × cue interaction ($F_{1,17} = 9.19$, $P = 0.008$), with post hoc contrasts showing lower entry rates during the degraded versus nondegraded cue during contingency degradation ($P < 0.05$) but not acquisition ($P > 0.05$). However, there was no training phase × cue × group interaction ($F_{1,17} = 0.06$, $P = 0.808$), suggesting that the optogenetic manipulation did not affect conditioned entries.

We also optogenetically inhibited DA release in the NAc during noncontingent rewards. VTA DA axons in the NAc expressing the synaptic terminal-inhibiting opsin, eOPN3, were targeted with a green light during noncontingent reward deliveries during the ITI (see Materials and Methods). This experiment did not include a Cre-negative control group, and instead, Cre-positive rats underwent 8 sessions of contingency degradation with optogenetic inhibition ("opto") followed by 11 sessions without inhibition ("post-opto"). Like the result of the previous experiment, optogenetic inhibition during noncontingent rewards did not affect cue-elicited entry rates (main effect of cue: $F_{1,6} = 18.34$, $P = 0.005$; no cue × training phase interaction: $F_{1,6} = 0.42$, $P = 0.540$; Fig. 5F).

Unlike eNpHR, for which there has been extensive documented use in VTA DA neurons (*22*–*25*), eOPN3 is a relatively new opsin (*26*). To confirm the functionality of eOPN3, rats were next trained to press levers for pellet rewards. One lever delivered pellets and occasionally triggered laser stimulation (see Materials and Methods), while the other lever only delivered pellets. Rats showed a deficit in learning to press the laser-paired lever (main effect of lever: $F_{1,6} = 7.09$, $P = 0.037$; Fig. 5G). This result suggests that the inhibitory opsin was functional.

Noncontingent VTA DA neuron activation attenuates a component of conditioned responding

The previous experiment shows that DA neuron activity elicited by noncontingent rewards is not necessary during contingency degradation to attenuate condition responding. We next asked whether DA neuron activity is sufficient to attenuate conditioned responding during a variation of contingency degradation, using DA neuron stimulation as the trial outcome. TH-Cre rats expressing channelrhodopsin (ChR2) in VTA DA neurons were first trained to associate a 7-s compound cue with unilateral VTA DA neuron optogenetic stimulation (1 s, 20 Hz, delivered at cue offset) for 12 daily sessions (Fig. 6B). TH–Cre-negative rats that did not express ChR2 were run in parallel. We previously reported that pairing localizable cues with unilateral VTA DA neuron stimulation results in conditioned locomotion in rats—specifically, approach toward the cue and full body rotation contralateral to the stimulated hemisphere (*14*). Therefore, we used DeepLabCut (*27*) to track rat body parts during this procedure to detect any possible changes in behavior during cue-DA stimulation pairings.

TH-Cre rats expressing ChR2 acquired a modest level of conditioned locomotion during Pavlovian acquisition, but not enough to render a group × session interaction (all tests, $F_{1,29} < 3.50$, $P > 0.073$; Fig. 6C). This modest level of responding was ideal for achieving a sub-asymptotic level of behavioral responding to avoid overtraining and was expected using this modification of our prior approach in which cues and stimulation do not overlap (*14*). During post-acquisition, rats were split into groups that were either maintained on the same conditioning protocol or switched to contingency degradation. During contingency degradation, cues were still followed by laser stimulation, but ITIs were filled with random deliveries of stimulation. The mean interval between stimulations during the ITI was designed to match the mean interval between cues and stimulations during trials.

Unexpectedly, the degree of conditioned locomotion, measured as head velocity, did not differ between maintained and degraded groups during post-acquisition (Fig. 6C). Repeated measures ANOVA's detected significant effects of group (all tests, $F_{1,26} > 4.20$, $P < 0.027$), but no main effects of session or group × session interactions. Post hoc contrasts on the baseline-subtracted data revealed that the two TH-Cre groups showed greater conditioned locomotion than the TH–Cre-negative group during the final conditioning session ($P < 0.05$), but the two TH-Cre groups did not differ from each other ($P > 0.05$).

Locomotion was measured by tracking the velocity of the rats' heads. There were other body parts that were tracked, and the angle,

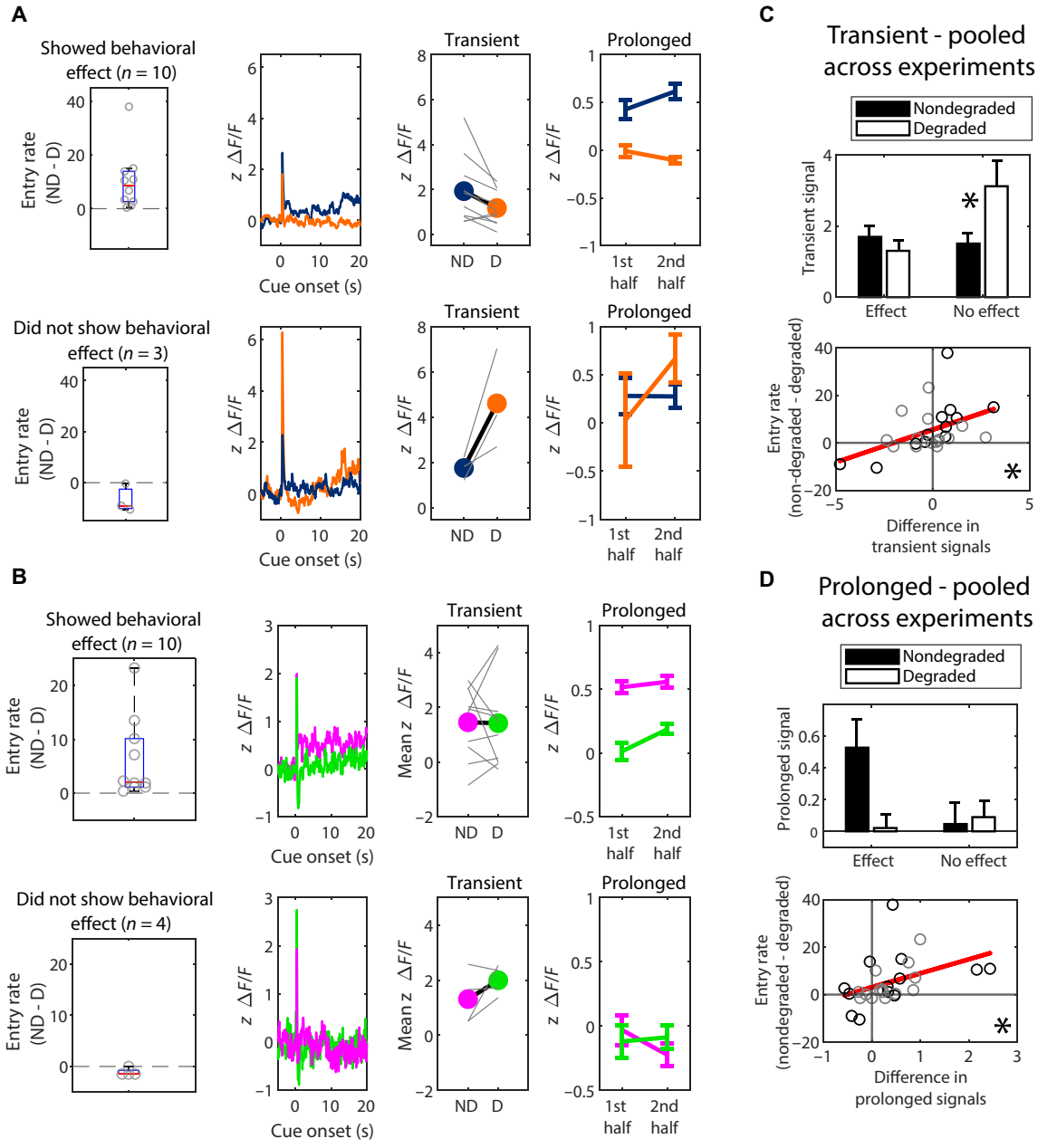

Fig. 4. Transient and prolonged components of cue-evoked DA activity and release predict conditioned responding during contingency degradation. (**A**) Data from the VTA GCaMP experiment broken down by rats that did and did not show behavioral sensitivity to contingency degradation. Photometry data only from the cue-outcome interval are shown (mean traces averaged across rats). ND, nondegraded; D, degraded. In the rightmost plots, the first and second halves refer to portions of the trial that the prolonged component spans. (**B**) Same as (A), but for the NAc dLight experiment. (**C**) Top: The transient response to cue onset (means ± SEM) was analyzed as a function of cue type and group. Bottom: The difference in the transient responses to degraded and nondegraded cues was used to predict behavior on a continuous measure. *$P < 0.05$. Black points come from the VTA GCaMP experiment and gray points are from the NAc dLight experiment. (**D**) Same as (C), but for the prolonged DA component. All data are from the last session of contingency degradation.

distance, and velocity between all pairs of points were computed for every frame. This created 135 different measurements per frame. The trial-averaged measures were compared between TH-Cre groups for the final session of post-acquisition, and paired t tests ($\alpha = 0.01$) revealed a significant group difference only for three measures: the distances between the middle of the tail and the left ear, right ear, and the middle of the head (Fig. 6D). Each measure showed the same group patterns, so only the latter measurement was analyzed (two-sample t test: $t_{20} = 2.88$, $P = 0.009$). To understand the meaning of this group difference, each video from the final session of post-acquisition was hand-scored. Rats were timed for the duration of rearing and rotating during and before each trial (Fig. 6E; see fig. S10 for data shown separately for trial and pretrial periods). An ANOVA revealed a group × behavior interaction ($F_{1,20} = 8.99$, $P = 0.007$).

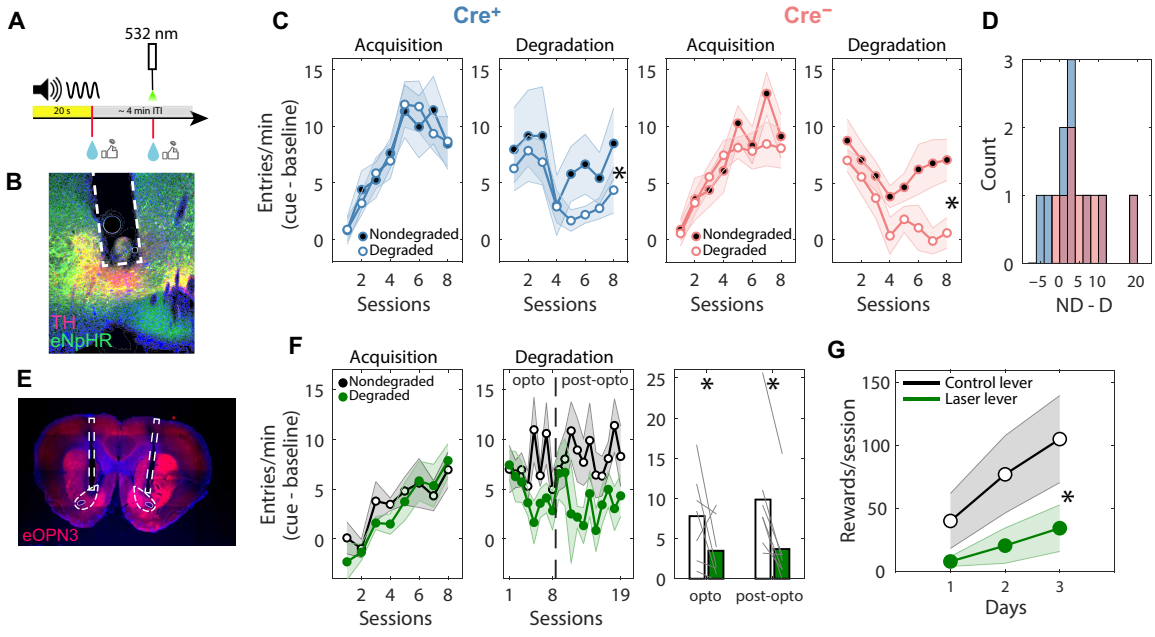

Fig. 5. Using optogenetic inhibition to test whether noncontingent dopamine is necessary to learn about contingency degradation. (**A**) Green laser onset occurred during the presentation of only noncontingent rewards during the ITI, but not contingent rewards at trial offset. The illustration shows only the degraded trial type, but nondegraded trials were also intermixed. (**B**) Coronal section from one TH-Cre rat injected with DIO-eNpHR in the VTA and implanted with bilateral fibers in the VTA. Dashed line outlines the fiber tract. Blue = 4′,6-diamidino-2-phenylindole (DAPI); red = TH; green = eNpHR. (**C**) Mean (±SEM) port entry rates during each cue, with baseline subtracted. *$P < 0.05$. Steel blue = Cre-positive; light coral = Cre-negative. (**D**) Distribution of the differences between entry rates during nondegraded (ND) and degraded (D) cues, shown for each group. (**E**) Coronal section from one TH-Cre rat injected with SIO-eOPN3 in the VTA and implanted with bilateral fibers in the NAc core. Dashed lines outline the fiber tracts. Blue = DAPI; red = eOPN3 terminals. (**F**) Mean (±SEM) port entry rates during the cue for each phase of the experiment. The vertical dashed line in the middle figures represents the point at which laser stimulation was removed from the experimental sessions. *$P < 0.05$. Gray lines in the right-most graph are individual rats. (**G**) Mean (±SEM) reward rates during the acquisition of lever pressing on an FR1 schedule for rats expressing eOPN3. *$P < 0.05$. Both levers were associated with pellet rewards, but only one was associated with green laser stimulation.

Post hoc contrasts showed that the group that underwent contingency degradation showed a lower rate of rotating compared to the control group ($P < 0.05$), while the rearing rates did not significantly differ ($P > 0.05$). Together, these results show that a cue directly paired with VTA DA activity will elicit conditioned rotation only when VTA DA activity is contingent on the cue.

The effects of contingency degradation on DA signals and behavior are better explained by ANCCR (adjusted net contingency for causal relations) than canonical TDRL

It is difficult to explain many of our empirical observations during contingency degradation from the point of view of a standard TDRL model. Simulations failed to show the sensitivity of either behavioral or dopaminergic responses to contingency degradation (fig. S11). We therefore simulated a recently published reinforcement learning model called ANCCR (adjusted net contingency for causal relations), which can successfully account for a range of phenomena related to Pavlovian conditioning and NAc DA release (*17*). The model builds representations of contingency between pairs of events by learning how often a given event predicts all other events (successor representation), and how often a given event is preceded by all other events (predecessor representation). These representations are combined to yield the net contingency (NC), and the DA response to an event is hypothesized to scale with the sum of all the event's net contingencies, plus the intrinsic meaningfulness of the event.

Model simulations reproduced the mean photometry transients aligned to cues and trial outcomes during outcome-selective contingency degradation (Fig. 7B). The same model parameters also generated a pattern of regression coefficients that were similar to the empirical coefficients when examining the history dependence of trial outcome responses (Figs. 2C and 3C), although the nondegraded coefficient one trial back failed to reach a negative value (Fig. 7C). However, the model did reproduce the results of the optogenetic experiments. The simulated behavioral response to the degraded cue was not affected by inhibiting the DA response during noncontingent rewards during the ITI (Fig. 7E), as in the experiment in which VTA DA neurons were inhibited using eNpHR (Fig. 5). In addition, like the experiment in which VTA DA neurons were stimulated using ChR2 (Fig. 6), cue-evoked responding was attenuated when noncontingent DA responses were simulated during the ITI (Fig. 7F). These results show that many of the DA responses during contingency degradation can be explained by the ANCCR model in which DA signals whether an event causes other meaningful events [(*17*), see Materials and Methods for computational algorithm].

DISCUSSION

We provide experimental evidence showing that DA neuron activity in the midbrain and DA release in the ventral striatum are sensitive to changes in the contingency between specific cue-reward associations. We also demonstrate that noncontingent DA neuron

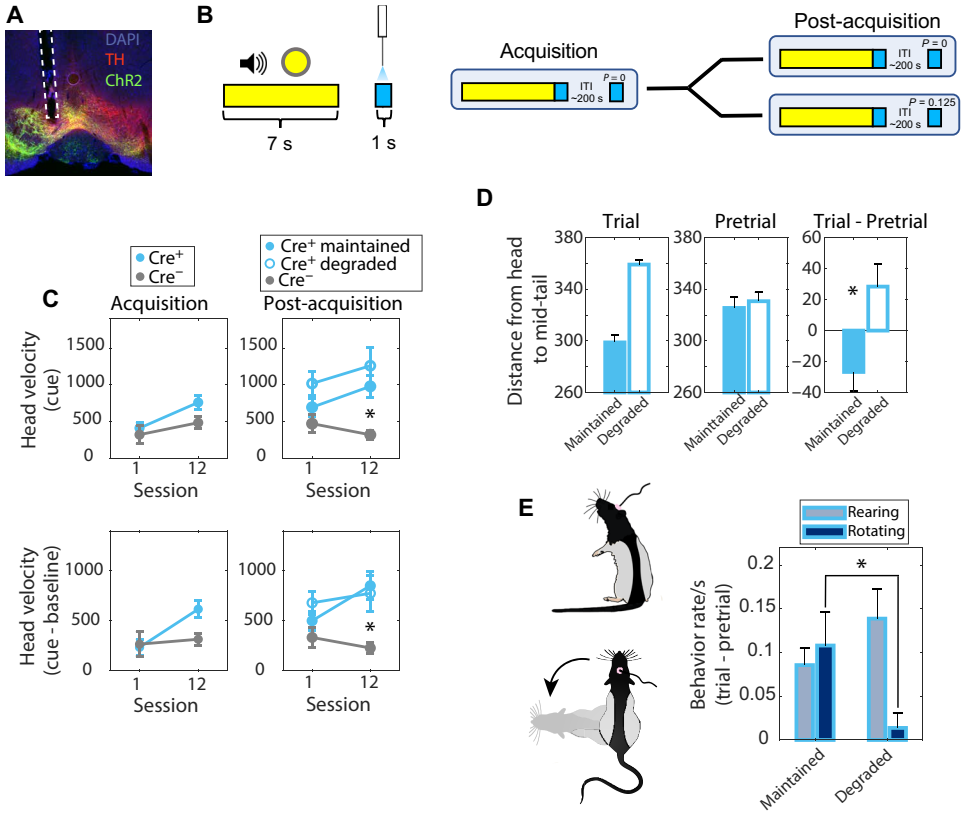

Fig. 6. Using optogenetic stimulation to manipulate the contingency between a cue and VTA DA neuron activity. (**A**) Coronal section from one TH-Cre rat injected with DIO-ChR2 in the VTA. Dashed line outlines the fiber tract. (**B**) Experiment design. During post-acquisition, some rats were maintained on the same conditioning protocol (top) or switched to contingency degradation (bottom). (**C**) Group velocity means (±SEM) during the cue period (top) and with baseline subtracted (bottom). (**D**) The group mean (±SEM) distances, in pixels, between the head and the middle of the tail during the trial (left), before the trial (middle), and as a difference between the trial and pretrial periods (right). *$P < 0.05$. (**E**) Left: Illustrations of rearing and rotating behavior. Right: Group mean (±SEM) rates of rearing and rotating during the trial period with baseline subtracted. *$P < 0.05$.

activity is sufficient, but not necessary, to block a conditioned response. These findings pose a substantial challenge to contiguity-based accounts of DA and learning, such as TDRL.

To reveal how cue-reward contingencies are reflected in DA dynamics, we used outcome-specific Pavlovian contingency degradation where animals learned two distinct cue-reward contingencies, after which one of the rewards was delivered during the ITI noncontingently. This manipulation degrades one cue-reward contingency while leaving the other intact (*1*, *3*), allowing a within-subject comparison of behavior and DA dynamics. We first showed that delivering rewards during the ITI attenuated the anticipatory cue-evoked port entries specifically for the cue that underwent contingency degradation without inducing specific satiety or context overshadowing. We next showed, using fiber photometry, that the cue- and reward-evoked DA responses were attenuated during the degraded trials, and the negative dip that characterizes the DA response to reward omission was less modulated during degraded trials. In addition, a trial-by-trial analysis showed that DA transients at the time of trial outcome were less sensitive to trial outcome history during degraded trials. Some of these findings agree with a prior study of NAc dLight recordings in mice before and after contingency degradation, which found that the cue-evoked DA response was suppressed, as it was in our rats (*17*). Notably, this study did not report impacts on the trial outcome response and did not use outcome-selective contingency degradation.

We then used optogenetics to ask how manipulations of DA neuronal activity and release influenced conditioned behavior during contingency learning. Bilateral optical inhibition did not change the course of behavior during contingency degradation—the gradual decrease in cue-evoked behavioral responding that typifies contingency degradation was maintained despite optically inhibiting DA neuronal activity and release at the time of noncontingent rewards. This finding suggests that the evaluative process involved in learning to suppress a previously acquired conditioned response can in some respect be independent of DA signaling. On the other hand, when cues were temporally paired with unilateral optical stimulation of DA neurons and then stimulation was also delivered during the ITI, cue-evoked conditioned rotations became less prevalent. This is reminiscent of an earlier finding where noncontingent optical stimulation of VTA DA terminals abolished latent inhibition (*28*). Together with the fiber photometry findings, these results show that the dynamics of mesostriatal DA change when the cue-reward contingency changes independent of contiguity, and, while the DA response evoked by noncontingent rewards is not necessary for learning contingency degradation, it is sufficient to attenuate conditioned responding at least when learning is completely DA-dependent.

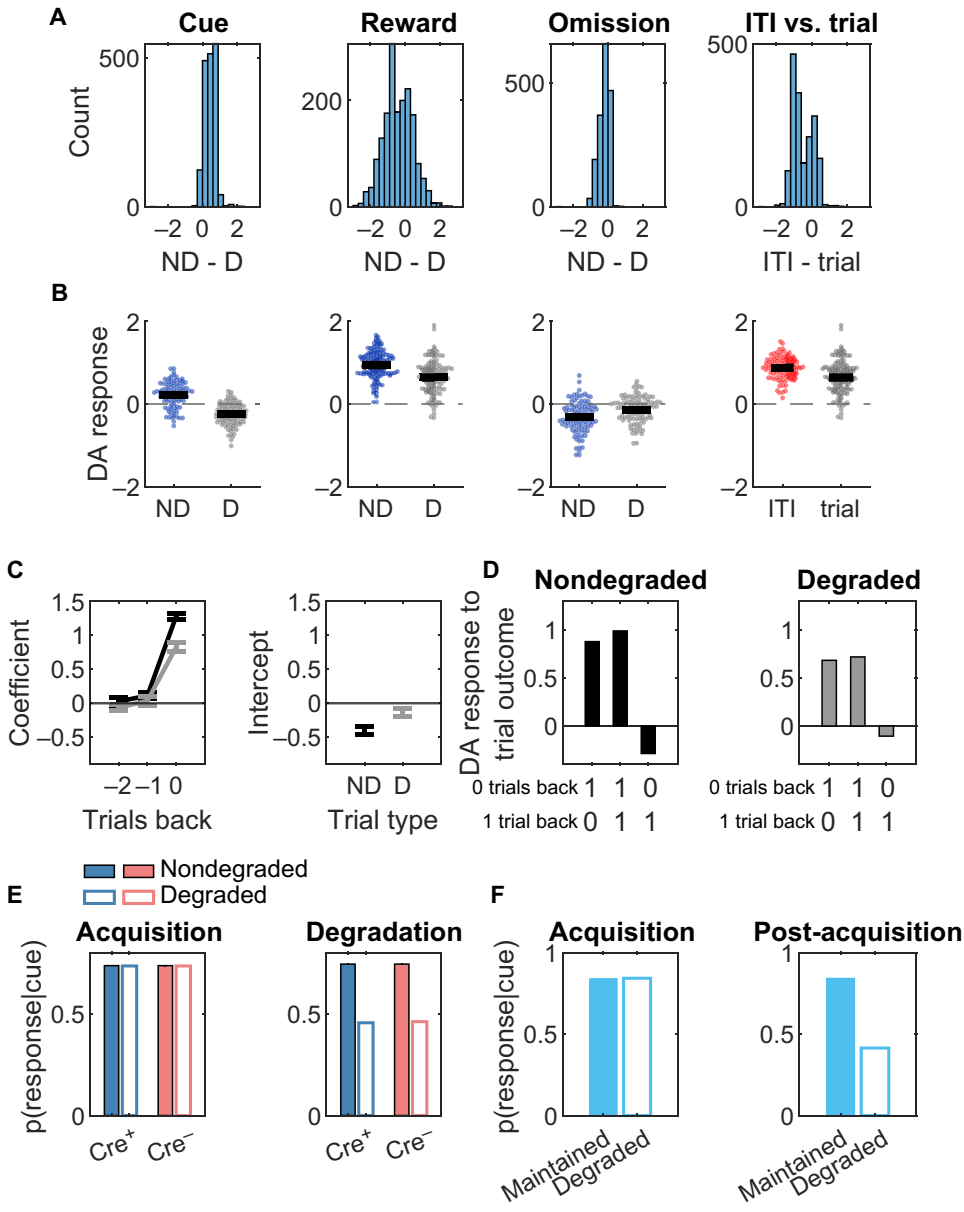

Fig. 7. ANCCR (adjusted net contingency for causal relations) simulations. (**A**) Distributions of simulated DA responses to trial events across all simulated parameter combinations. ITI = degraded reward presented during the ITI; trial = degraded reward presented at the end of a trial. (**B**) DA responses to trial events from the best-fitting parameter combination. Means are taken from the final 50 trials of contingency degradation. (**C**) Mean (±SEM) coefficients and intercepts when regressing the DA response to the trial outcome on trial outcome history. Means are taken from the final 50 trials of degradation. (**D**) Predicted DA response at the time of the outcome on trial *n* as a function of the outcome on trial *n* and trial *n* − 1. 1 indicates reward and 0 indicates omission. Means are taken from the final 50 trials of degradation. (**E**) Simulated behavioral response to the degraded cue when DA response to noncontingent rewards is inhibited (Cre-positive) or not (Cre-negative). (**F**) Behavioral simulation results when a DA response follows a 7-s cue (acquisition) and is then also delivered noncontingently during the ITI (post-acquisition, degraded) or not (post-acquisition, maintained).

It is difficult to account for these results within the framework of TDRL. During contingency degradation, the values of the cues remain unchanged because they are followed by the same rewards at the same intervals and with the same probabilities as before contingency degradation. If phasic DA is to be equated with the TDRL prediction error, but the value remains unchanged during contingency degradation, then DA dynamics should not differ between degraded and nondegraded trials because the TDRL prediction error is the first derivative of the value function (*7*). In contrast, ANCCR was able to capture almost all the major results (see also Supplementary Text).

The present dataset is not the first to challenge TDRL as an explanation for the DA function. Experiments in rats have shown that VTA DA neurons mimic a prediction error that accounts for expectations concerning stimulus identity, not just value (*12*, *25*, *29*). These findings led to the proposal that VTA DA neurons compute sensory prediction errors as part of the successor representation algorithm, which learns the frequencies with which a given state is

followed by every other state (*30*). This learning algorithm is similar to TDRL, but it learns stimulus-stimulus associations and reward identities. ANCCR takes this idea a few steps further by quantifying a way for agents to also learn how frequently a given state is preceded by every other state and accounting for the base rate of that state—known as the predecessor representation contingency (PRC). The successor and predecessor representation contingencies are then combined to yield the NC between every pair of states. Our simulations show that, while the successor representation contingencies between the two cue-reward pairs were similar on average, the PRC was high at the time of reward following the nondegraded cue and low during reward following the degraded cue (fig. S12).

These proposed underlying representations also help to explain why the DA response to reward following the degraded cue is smaller than to reward following the nondegraded cue. The ANCRR model considers DA to scale with the sum of net contingencies between a given event and all meaningful causal targets. During outcome-selective contingency degradation, the DA response to each reward type will depend on its contingency with respect to all other events. The predecessor representation contingencies between degraded rewards (i.e., the reward type delivered during the ITI) and all other events are negative (fig. S12) because its base rate is high, and the PRC is negatively affected by a high base rate (see Materials and Methods for computational algorithm).

There are two instances where ANCCR did not fully capture some features of our dataset. First, using the set of parameters that faithfully reproduce the empirical photometry results, ANCCR fails to accurately reproduce the pattern of regression coefficients that we observe when modeling the DA response to the trial outcome as a function of local reward history. Specifically, the coefficient associated with the nondegraded trial type one trial back in time goes to zero rather than negative. This negative coefficient one trial back is important for explaining some of the similarities between DA dynamics and reward prediction errors because it reflects the inhibition of DA neurons during positive reward expectation (*19*, *31*). Second, in only a minority of simulated cases (20%) did ANCCR predict a larger DA response to a reward delivered during the ITI than at the end of a trial (Fig. 7A). Key to reproducing this effect was a small learning rate for updating the baseline average rate of events (k; see Materials and Methods for computational algorithm). This was not sufficient, however, and the interactions among the model parameters were complex. Together, these results show that the predictions of ANCCR are highly parameter-dependent, but our results may be able to constrain the parameter space.

In summary, the main finding of this study is that DA responses to cues and rewards are sensitive to the contingency between them; mere temporal contiguity is not sufficient to elicit the prototypical DA responses to these events. In addition, DA signaling during noncontingent rewards was not necessary for rats to learn outcome-selective contingency degradation. The gradual decrease in cue-evoked behavioral responding that typifies contingency degradation was maintained despite optically inhibiting DA neuronal activity and release at the time of noncontingent rewards. Last, a simple TDRL model was insufficient to fully explain behavior and DA dynamics when contingency was manipulated. Instead, an alternative model, ANCCR, based on estimates of causality explains most aspects of the dataset. Note that the TDRL simulations performed in this study do not incorporate recent modifications that attempt to account for feature-specific prediction errors (*32*,*33*) (see "note added in proof"). These TDRL models postulate that prediction errors are vectorized to reflect a biased integration of separate streams of feature-specific predictions. It is therefore possible that individual DA neurons compute feature- and outcome-specific prediction errors during outcome-specific contingency degradation. It is uncertain whether the summation of such feature-specific prediction errors would resemble the photometry traces in the present report. The use of outcome-specific contingency degradation combined with single-unit DA neuron recordings will be a productive line of research going forward. The present work adds to a growing complement of studies that require us to expand our understanding of DA's role in learning (*12*, *14*, *25*, *34*–*38*).

MATERIALS AND METHODS

Rats and surgeries

All experimental procedures were performed in strict accordance with protocols approved by the Animal Care and Use Committee at Johns Hopkins University. Thirteen TH-Cre rats were used for fiber photometry recordings in the VTA (six females, seven males). Rats underwent surgery 4 to 5 weeks before the beginning of the behavioral experiment. During surgery, a virus [1 μl of AAVDJ-EF1a-DIO-GCaMP6f diluted to 5×10^{12} particles/ml in phosphate-buffered saline (PBS)] was injected at the following coordinate relative to bregma: AP −5.8, ML +0.7, DV −8. An optic fiber (Doric Lenses) attached to an adapter was then lowered to 0.2 mm above the virus coordinate.

Fourteen wild-type Long-Evans rats were used for fiber photometry in the NAc (six females, eight males). Rats underwent surgery 4 to 5 weeks before the beginning of the behavioral experiment. During surgery, a virus (1 μl of AAV5-hSyn-dLight1.2 diluted to 4.6×10^{12} particles/ml in PBS) was injected at the following coordinate relative to bregma: AP +1.75, ML +1.7, DV −7. An optic fiber (Doric Lenses) attached to an adapter was then lowered to 0.2 mm above the virus coordinate.

Eight Long-Evans rats were used for the context switch and food preference experiments (four males, four females). These rats did not undergo surgery.

Nineteen Long-Evans rats (10 TH-Cre, 9 TH–Cre-negative) were used for the optogenetic inhibition experiment in the VTA (9 females, 10 males). Rats underwent virus injection and fiber implant surgeries 4 weeks before the beginning of the behavioral experiment. During surgery, a virus (AAV5- Ef1α -DIO-eNPHR3.0-eYFP diluted to 4×10^{12} particles/ml in PBS) was injected bilaterally in the VTA at the following coordinates relative to bregma: 1 μl at AP −5.8, ML ±0.7, DV −8.4; 0.5 μl at AP −5.8, ML ±0.7, DV −7.4. Optic fibers attached to ferrules were then lowered bilaterally to the VTA at the following coordinates at a 15° angle: AP −5.8, ML ±2.71, DV −7.76.

Seven TH-Cre rats were used for the optogenetic inhibition experiment in the NAc (three females, four males). Rats underwent virus injection surgeries 11 weeks before the beginning of the behavioral experiment. During surgery, a virus (AAV1-hSyn1-SIO-eOPN3-mScarlet-WPRE diluted to 5×10^{12} particles/ml in PBS) was injected bilaterally in the VTA at the following coordinates relative to bregma: 1 μl at AP −6.2, ML ±0.7, DV −8.4; 0.5 μl at AP −6.2, ML ±0.7, DV −7.4; 1 μl at AP −5.4, ML ±0.7, DV −8.4; 0.5 μl at AP −5.4, ML ±0.7, DV −7.4. Eight weeks later, optic fibers attached to ferrules were then lowered bilaterally to the NAc at the following coordinates at a 10° angle: AP +1.75, ML ±2.96, DV −6.7.

Twenty-nine rats were used for optogenetic stimulation in the VTA (17 females, 12 males). Rats underwent surgery 4 to 5 weeks before the beginning of the behavioral experiment. During surgery, a virus (AAV5-Ef1a-DIO-ChR2-eYFP diluted to 4.2×10^{12} particles/ml in PBS) was injected at the following coordinates relative to bregma: 1 μl at AP −6.2, ML +0.7, DV −8.4; 0.5 μl at AP −6.2, ML +0.7, DV −7.4; 1 μl at AP −5.4, ML +0.7, DV −8.4; 0.5 μl at AP −5.4, ML +0.7, DV −7.4. An optic fiber attached to a ferrule was then lowered to the following coordinates: AP −5.8, ML +0.7, DV −7.5. All implants were secured to the skull with dental acrylic applied around skull screws, the base of the ferrule, and, in rats undergoing photometry recordings, the adapter.

Pavlovian conditioning with food rewards

Pavlovian conditioning experiments with food rewards were conducted in plexiglass chambers with grid floors surrounded by a sound-attenuating cubicle (Med Associates). Chambers were equipped with a food reward port that contained an infrared beam. Each time the beam was broken, a port entry was recorded.

Rats were food deprived to 85% of their ad libitum weight. Experiments began with two sessions of port training during which one type of food reward (45 mg of grain pellet or 0.1 ml of 20% sucrose) was delivered randomly into a food port every 60 s on average. Pellet deliveries were associated with the sound of the dispenser and a clinking of the pellet into the port, and sucrose deliveries were associated with the sound of the syringe pump and two clicks separated by 0.2 s. Each port training session contained different food rewards, but the same port was used for both reward types.

During the acquisition phase of Pavlovian conditioning, each food reward was preceded by a unique auditory cue (pure tone or white noise). When an auditory cue was presented, it lasted for 20 s and was followed by a 0.5 probability of its associated reward. The ITI was drawn from an exponential distribution with a mean of 4 min and a range of 21 s and 14 min. A session lasted 70 min, and within a session, each trial type was presented eight times in random order. Although the reward was probabilistic, rats were guaranteed to experience four rewarded trials and four nonrewarded trials of each type. Pavlovian acquisition lasted for eight sessions.

During contingency degradation, the task contingencies remained the same except one type of food reward was delivered during the ITI noncontingently. Specifically, the reward was delivered during the ITI every 20 s with 0.5 probability except when the upcoming trial was less than 20 s away. This was to avoid noncontingent rewards being delivered during the 20-s pre-cue baseline. Contingency degradation lasted for eight sessions. Cue-reward assignments, noncontingent reward identity, and sex were counterbalanced. Following contingency degradation, one session of reacquisition was given during which noncontingent rewards were no longer delivered during the ITI.

Context switch

Rats were put through the same series of Pavlovian conditioning and contingency degradation sessions as described above. Following the last session of degradation, two extinction sessions were conducted during which four nonrewarded tone and noise trials were presented in random order and at random times (exponentially distributed ITI with a mean of 4 min). One session was conducted in the normal context and the other in a modified context, with an additional session of contingency degradation separating the two tests. The modified context had a floor made from plastic with raised points in a honeycomb pattern, stripe-patterned walls, and a lemon scent. The normal context had grid floors without the stripped patterns or any special scent. Cue-reward assignments, noncontingent reward identity, and sex were counterbalanced. The order of testing was counterbalanced with cue-reward assignments and noncontingent reward identity.

Food preference tests

Following the context tests, two food preference tests were conducted during which rats were free to consume the same pellet and sucrose rewards used in the Pavlovian conditioning experiment. The first test was conducted immediately after an additional session of contingency degradation. The second test was conducted the following day without any prior experimental session. Tests lasted for 30 min. The weights of the pellets and sucrose were recorded before and after the tests.

Fiber photometry

A fluorescence mini-cube (Doric Lenses) transmitted light streams from a 465-nm light-emitting diode (LED) sinusoidally modulated at 330 Hz that passed through a green fluorescent protein (GFP) excitation filter, and a 405-nm LED modulated at 120 Hz. Both 465- and 405-nm streams were band-pass–filtered. LED power was set at 28 μW for the 405 stream and 68 μW for the 465 stream. GCaMP6f and dLight1.2 fluorescence from neurons below the fiber tip in the brain was transmitted via this same low-autofluorescence fiber cable (400 nm, 0.52 NA) back to the mini-cube, where it was passed through a GFP emission filter, amplified, and focused onto a high sensitivity photoreceiver (Doric Lenses). A real-time signal processor (RZ5P, Tucker-Davis Technologies) running Synapse software modulated the output of each LED and recorded photometry signals, which were sampled from the photodetector at 6 kHz. The signals generated by the two LEDs were demodulated and decimated to 1020 Hz for recording to disk.

For analysis, signals were downsampled to 102 Hz, and the 465-nm signal was normalized to the 405-nm signal by computing $\Delta F/F$. Specifically, the best-fitting line relating the 465-nm signal to the 405-nm signal was estimated, and the 405-nm signal was then transformed by multiplying by the regression slope and adding on the y intercept. This puts the 405-nm data in the range of the 465-nm data. $\Delta F/F$ was computed as (465 nm − transformed 405 nm)/(transformed 405 nm).

Optogenetic inhibition of VTA DA neurons during Pavlovian contingency degradation

Rats with optic fibers targeting the VTA were put through the identical experiment described above (see the "Pavlovian conditioning with food rewards" section). During contingency degradation, rats received green laser stimulation delivering 4 s of constant 532-nm light (12 to 15 mW) at the onset of every noncontingent reward delivered during the ITI. Rats were tethered to a bilateral patch cord during every session of the experiment.

Optogenetic inhibition of DA terminals in NAc during Pavlovian contingency degradation

Rats with optic fibers targeting eOPN3-expressing VTA axons in the NAc were put through a similar behavioral experiment described above (see the "Pavlovian conditioning with food rewards" section). During contingency degradation, rats received green laser stimulation

delivering 1 s of constant 532-nm light (10 mW) at the onset of the first reward delivered during the ITI. The effect of eOPN3 activation lasts for minutes (*26*), and this creates a trade-off between minimizing the eOPN3 activation that continues into the next trial and the number of rewards during the ITI that are given in the absence of eOPN3 activation. To strike a reasonable balance between this trade-off, the range of the ITIs was changed to 2 to 6.5 min. The mean ITI was maintained at 4 min. In addition, no rewards were delivered during the ITI when the upcoming trial was less than 60 s away. Rats were tethered to a bilateral patch cord during every session of the experiment.

Optogenetic inhibition of DA terminals in NAc during instrumental acquisition

Rats that underwent inhibition of DA terminals in the NAc during Pavlovian contingency degradation were next trained to press two levers on fixed-ratio 1 schedules in separate sessions. Two 30-min sessions were run each day for a total of 3 days during which one lever was presented and rats were free to press the lever for a grain pellet reward at any time. One lever was assigned as the laser lever and the other was assigned as the control lever. When the laser lever was pressed, a pellet reward was triggered along with 1 s of constant 532-nm light (10 mW). The laser was withheld if the previous laser onset occurred within the last 60 s. The control lever session did not include any laser stimulation, although rats were still tethered to the patch cord. The order of testing was balanced across rats and alternated each day.

Pavlovian conditioning with optogenetic stimulation

Rats with fibers targeting ChR2-expressing neurons in the VTA were tethered to a patch cord connected to a laser via a commutator. Lasers delivered 473-nm blue light (12 mW). Conditioning was divided into two phases: acquisition and post-acquisition. During acquisition, a compound cue (panel light and pure tone), was presented for 7 s followed by 1-s laser stimulation (20 5-ms pulses at 20 Hz). ITIs were drawn from an exponential distribution with a mean of 200 s and a range of 15 to 615 s. Sessions lasted for 86 min during which 25 trials were presented. Rats were split into two groups during acquisition: Cre-negative (four females, three males) and Cre-positive (14 females, 10 males).

During post-acquisition, some rats were maintained on the same conditioning protocol (Cre-positive: seven females, three males; Cre-negative: one female, two males), while others underwent contingency degradation (Cre-positive: six females, six males; Cre-negative: three females, one male). All experiment parameters remained identical except noncontingent 1-s laser stimulations were delivered during the ITI with a probability of 0.125 every second. Two Cre-positive rats from the acquisition phase were dropped from post-acquisition analysis because one became disconnected from the patch cords during the last session and the other scratched its head to the point of bleeding before the start of the last session.

Histology

At the end of all experiments, rats were perfused transcardially with 0.9% saline followed by 4% paraformaldehyde (PFA). Brains were removed and stored in PFA for 1 hour followed by 30% sucrose in PBS for 72 hours. Coronal sections (50 μm thick) were cut using a cryostat, and sections were stored in PBS at 4°C. Brain sections were mounted on microscope slides, coverslipped with VECTASHIELD Antifade Mounting Medium with 4′,6-diamidino-2-phenylindole (DAPI), and examined with a fluorescent microscope (Zeiss).

Statistical analysis

Statistical tests on summary data included repeated measures ANOVA and *t* tests with a significance threshold of 0.05 except where noted in the main text. Significant interactions were followed up with post hoc contrasts using the recommendations of Rodger (*38*).

All $\Delta F/F$ photometry traces were *z*-scored relative to a pretrial 20-s baseline. When *z*-scoring noncontingent rewards during the ITI, the nearest pretrial baseline was used. To summarize photometry responses to cues, the *z*-scored $\Delta F/F$ was averaged starting from the time of cue onset and ending 5 s later. To divide cue-evoked responses into transient and prolonged components, the transient beginning and end periods were defined as the times after cue onset at which the lower bound of a 95% confidence interval passed above and dropped back down to 0, respectively (*40*). For the VTA GCaMP6F experiment, this interval was between 255 and 549 ms after cue onset. For the NAc dLight experiment, this interval was between 245 and 372 ms after cue onset. The prolonged component of the cue-evoked responses was the mean of the photometry signal starting from the end of the transient window and ending right before cue offset. To summarize photometry responses to rewards, the *z*-scored $\Delta F/F$ was averaged starting from the time of reward onset and ending 10 or 5 s later for VTA GCaMP6f and NAc dLight, respectively.

Reward omission responses were divided into positive and negative phases. The positive phase was averaged starting from the moment after cue offset and ending when activity significantly dropped below baseline. The negative phase was averaged starting from the moment after cue offset when activity significantly dropped below baseline and ending when activity rose back up to baseline. To identify when the signal significantly dropped below baseline, the time at which the upper bound of a 95% confidence interval dropped below 0 was used [Jean-Richard-dit-Bressel, Clifford, & McNally (*40*)]. For the VTA GCaMp6f experiment, the averaging window for the negative phase was between 3.31 and 9.76 s after the trial. For the NAc dLight experiment, the window was between 0.66 and 2.29 s after the trial.

To quantify how local reward history influenced conditioned responding and photometry responses to cues and outcomes, we used multiple regression. The number of cued port entries on trial *t* was modeled as a linear combination of trial outcomes

$$\text{entries}(t) = \beta_0 + \beta_1 * R(t-1) + \beta_2 * R(t-2) \quad (1)$$

where *R* is the trial outcome (1 if rewarded, 0 if omitted). The GCaMP and dLight responses to the trial outcome and the cue were modeled in a similar way

$$d(\text{outcome}_t) = \beta_0 + \beta_1 * R(t) + \beta_2 * R(t-1) + \beta_3 * R(t-2) \quad (2)$$

$$d(\text{cue}_t) = \beta_0 + \beta_1 * R(t-1) + \beta_2 * R(t-2) \quad (3)$$

The trial outcome was averaged over the start and end points used to define the reward omission response window (see previous paragraph). The number of past trial outcomes used in Eqs. 1 to 3 was chosen with consideration toward the amount of data available to train the regression models.

To predict conditioned behavior from photometry signals on a trial-by-trial basis, three multiple regression analyses used the mean z-scored $\Delta F/F$ response to cue and trial outcome, as well as the trial number

$$\text{entry timing}(t) = \beta_0 + \beta_1 * d(\text{cue}_t) + \beta_2 * d(\text{outcome}_{t-1}) + \beta_3 * \text{trial}_t \quad (4)$$

$$\text{entry rate}(t) = \beta_0 + \beta_1 * d(\text{cue}_t) + \beta_2 * d(\text{outcome}_{t-1}) + \beta_3 * \text{trial}_t \quad (5)$$

$$\text{entry latency}(t) = \beta_0 + \beta_1 * d(\text{cue}_t) + \beta_2 * d(\text{outcome}_{t-1}) + \beta_3 * \text{trial}_t \quad (6)$$

Entry rate was defined as the total number of entries during the 20-s cue. Entry timing was defined as the area under the curve of the cumulative proportion of entries during the 20-s cue. Entry latency was defined as the time from cue onset to the first entry during the 20-s cue.

Behavioral data during conditioning with optogenetic stimulation were derived from body part location estimates using DeepLabCut (*27*). Each video frame contained the estimated x-y positions of the following body parts and environmental cues: left ear, right ear, middle of the head between the ears, middle of the back, base of the tail, middle of the tail, four corners of the behavioral chamber, and the left panel light, which functioned as part of the compound cue. DeepLabCut also generated a confidence measure ranging from 0 to 1, and trials containing frames with confidence measures less than 0.95 were excluded from analyses. To derive the velocity of the rat, the distance between the middle of the ears and the panel light was measured in pixels for each frame, and the differences between the frame-by-frame distances were computed. Video scoring was performed by using a stopwatch to quantify the duration of rearing and rotating during each trial.

ANCCR simulations

All simulations were performed in MATLAB with the help of functions available on the Namboodiri lab GitHub (https://github.com/namboodirilab/ANCCR). Simulated experimental events mimicked the same contingencies during acquisition and contingency degradation, except the number of trials was increased to 500 per cue in each phase. We first simulated ANNCR on the photometry experiments, using 20 iterations per parameter combination. There were six free parameters set to the following ranges: T ratio = 0.2 to 1.4, $\alpha = 0.01$ to 0.3, $k = 0.01$ to 0.6, $w = 0.1$ to 0.7, threshold = 0.1 to 0.7, $\alpha_R = 0.1$ to 0.3. The winning parameter combination was determined by maximizing the correlation between the rankings of the following empirical and simulated means: ITI reward, degraded reward, nondegraded cue, and degraded cue. Once the winning parameter combination was identified, the model was simulated again for 100 iterations.

The winning set of parameters was then used to simulate behavioral responding during the optogenetic inhibition and stimulation experiments using 100 iterations per experiment.

ANCCR computations are described below. Event i is kept in memory across time steps according to an eligibility trace

$$E_{\leftarrow i} = \sum_{t_i \le t} e^{-\left(\frac{t-t_i}{T}\right)} \quad (7)$$

where T is the temporal decay parameter. T was always considered a fraction of the mean inter-reward interval during acquisition, and this fraction (T ratio) was a free parameter used to fit the model.

ANCCR involves computing the adjusted NC between all pairs of events and using that to generate DA and behavioral responses. The adjusted NC is derived from the NC, which is derived from the predecessor and successor representation contingencies. The predecessor representation between events i and j is updated at the time of j as

$$M_{\leftarrow ij} \equiv M_{\leftarrow ij} + \alpha[E_{\leftarrow ij} - M_{\leftarrow ij}] \quad (8)$$

where $E_{\leftarrow ij}$ measures the eligibility trace of i and the current time of j and $\equiv$ denotes an update operation. The predecessor representation quantifies how often the event i precedes the event j. The agent also keeps track of the baseline rate of all events according to

$$M_{\leftarrow i} \equiv M_{\leftarrow i} + k\alpha[E_{\leftarrow i} - M_{\leftarrow i-}] \quad (9)$$

where – represents random moments. We assumed it is updated every 0.2 s. The PRC is then calculated as

$$\text{PRC}_{\leftarrow ij} = M_{\leftarrow ij} - M_{\leftarrow i-} \quad (10)$$

The successor representation contingency (SRC) is then calculated using Bayes' rule and is updated as the time of j

$$\text{SRC}_{\rightarrow ij} = \text{PRC}_{\leftarrow ij} \frac{M_{\leftarrow j-}}{M_{\leftarrow i-}} \quad (11)$$

The successor representation quantifies how often the event j follows the event i. The predecessor and successor representation contingencies were then combined into a weighted sum to yield the NC

$$\text{NC}_{\leftrightarrow ij} = w\text{SRC}_{\rightarrow ij} + (1-w)\text{PRC}_{\leftarrow ij} \quad (12)$$

The adjusted NC (ANCCR) between i and j is then calculated to account for possible causes of i, like when it is consistently preceded by k

$$\text{ANNCR}_{\leftrightarrow ij} = \text{NC}_{\leftrightarrow ij} R_{ij} - \sum_{k \ne i} (\text{ANCCR}_{\leftrightarrow kj} \Delta_{k \leftarrow i}) \quad (13)$$

Here, Δ measures the recency of k with respect to i and is defined as

$$\Delta_{k \leftarrow i} = e^{-\left(\frac{t_j - t_k}{T}\right)} \quad (14)$$

R_{ij} is a causal weight that is updated according to

$$R_{ij} \equiv R_{ij} + \alpha_R \delta_{ij} \quad (15)$$

where δ_{ij} is a prediction error that is computed according to

$$\delta_{ij} = \begin{cases} R_{jj} - R_{ij}, & \text{DA}_j \ge 0 \\ (0 - R_{ij}) \dfrac{n_i^{-1} \Delta_{i \leftarrow j}}{n_k^{-1} \Delta_{k \leftarrow j}}, & \text{DA}_j < 0 \end{cases} \quad (16)$$

where R_{jj} is the externally signaled reward magnitude of j and n_i is the total number of times event i has been experienced. The DA response to an event i is the sum of the learned meaningfulness of stimulus i (ANCCRs of i with respect to all meaningful causal targets j) and the innate meaningfulness (b_i)

$$\text{DA}_i = \sum_j \text{ANCCR}_{\leftrightarrow ij} + b_i \quad (17)$$

In the original proposal (Eq. 17), the DA response was defined as the learned meaningfulness of a stimulus, but it has been revised to account for both the learned and innate meaningfulness of a stimulus in a subsequent paper [refer to (*34*) for the rationale behind this model update].

For j to be considered a meaningful causal target, the DA response must pass a threshold

$$\mathbb{I}(j \in \mathrm{MCT}) = \begin{cases} 1, & \text{if } \mathrm{DA}_j + b_j > \theta \\ 0, & \text{otherwise} \end{cases} \quad (18)$$

The variable b_j represents the innate meaningfulness of stimulus j and was always set to 0.5 for rewards and 0 for all other stimuli. To generate behavioral response probabilities, we applied a softmax function

$$p(\text{action} \mid \text{cue}_i) = \frac{e^{\beta V_i}}{\sum_j e^{\beta V_j}} \quad (19)$$

where β is the inverse temperature and was set to 5. V is the value of a cue and was defined as $\mathrm{NC}_{\leftrightarrow ij}\, R_{ij}$ − cost. The cost was set to 0.3.

The set of parameters that best fit the photometry data was T ratio = 1, α = 0.2, k = 0.01, w = 0.4, threshold = 0.7, α_R = 0.1.

TDRL simulations

We simulated TDRL in the same experiments as ANCCR. Event i is kept in memory across time steps t according to an eligibility trace

$$e(i)_{t+1} = \gamma\lambda e(i)_t + x(i)_t \quad (20)$$

where $x(i)_t$ is the activation of state i associated with event x at time t, and γ and λ are discount and eligibility decay parameters, respectively. The representation of events as discrete nonoverlapping states is known as the complete serial compound (*11*). The values of each state at time t are a weighted sum of their activations at time t

$$V_t(x) = w_t^T x_t = \sum_{i=1}^{m} w_t(i) x_t(i) \quad (21)$$

Weights are updated according to the size of the temporal difference prediction error, shown in brackets

$$w(i)_{t+1} = w(i)_t + \alpha \bullet [r_t + \gamma V_t(x_t) - V_t(x_{t-1})] \bullet e(i)_t \quad (22)$$

Under this model, the DA response to any given event is equal to the prediction error at the time of the event. Behavioral response probabilities were generated according to

$$p(\text{action} \mid \text{cue}_i) = \frac{e^{\beta V_i}}{\sum_j e^{\beta V_j}} \quad (23)$$

where β is the inverse temperature and was set to 5. V is the value of a cue as defined above $[V_t(x)]$. The cost was set to 0.3. γ was set to 0.85 and λ was set to 0.

Note added in proof: After this paper was accepted for publication, the authors requested to acknowledge a recent relevant paper: L. Qian, M. Burrell, J. A. Hennig, S. Matias, V. N. Murthy, S. J. Gershman, N. Uchida, The role of prospective contingency in the control of behavior and dopamine signals during associative learning. *bioRxiv* (2024). https://www.biorxiv.org/content/10.1101/2024.02.05.578961v1.

Supplementary Materials

This PDF file includes:

Supplementary Text
Figs. S1 to S12
References

REFERENCES AND NOTES

1. A. R. Delamater, Outcome-selective effects of intertrial reinforcement in a Pavlovian appetitive conditioning paradigm with rats. *Anim. Learn. Behav.* **23**, 31–39 (1995).
2. P. J. Durlach, D. O. Shane, The effect of intertrial food presentations on anticipatory goal-tracking in the rat. *Q. J. Exp. Psychol. B.* **46**, 289–318 (1993).
3. S. B. Ostlund, B. W. Balleine, Orbitofrontal cortex mediates outcome encoding in Pavlovian but not instrumental conditioning. *J. Neurosci.* **27**, 4819–4825 (2007).
4. R. A. Rescorla, Probability of shock in the presence and absence of CS in fear conditioning. *J. Comp. Physiol. Psychol.* **66**, 1–5 (1968).
5. C. R. Gallistel, P. D. Balsam, Time to rethink the neural mechanisms of learning and memory. *Neurobiol. Learn. Mem.* **108**, 136–144 (2014).
6. C. R. Gallistel, A. R. Craig, T. A. Shahan, Temporal contingency. *Behav. Processes* **101**, 89–96 (2014).
7. R. S. Sutton, A. G. Barto, *Reinforcement Learning: An Introduction* (MIT Press, 2018).
8. 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).
9. H. G. 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).
10. M. R. Roesch, D. J. Calu, G. Schoenbaum, Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. *Nat. Neurosci.* **10**, 1615–1624 (2007).
11. W. Schultz, P. Dayan, P. R. Montague, A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997).
12. R. Keiflin, H. J. Pribut, N. B. Shah, P. H. Janak, Ventral tegmental dopamine neurons participate in reward identity predictions. *Curr. Biol.* **29**, 93–103.e3 (2019).
13. 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).
14. 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).
15. 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).
16. R. S. Sutton, A. G. Barto, Time-derivative models of Pavlovian reinforcement, in *Learning and Computational Neuroscience: Foundations of Adaptive Networks*, M. Gabriel, J. Moore, Eds. (MIT Press, 1990), pp. 497–537.
17. 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. *Science* **34**, 642–685 (2022).
18. P. D. Balsam, M. R. Drew, C. R. Gallistel, Time and associative learning. *Comp. Cogn. Behav. Rev.* **5**, 1–22 (2010).
19. H. M. Bayer, P. W. Glimcher, Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron* **47**, 129–141 (2005).
20. D. J. Ottenheimer, B. A. Bari, E. Sutlief, K. M. Fraser, T. H. Kim, J. M. Richard, J. Y. Cohen, P. H. Janak, A quantitative reward prediction error signal in the ventral pallidum. *Nat. Neurosci.* **23**, 1267–1276 (2020).
21. 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).
22. 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).
23. R. Luo, A. Uematsu, A. Weitemier, L. Aquili, J. Koivumaa, T. J. McHugh, J. P. Johansen, A dopaminergic switch for fear to safety transitions. *Nat. Commun.* **9**, 2483 (2018).
24. J. E. McCutcheon, J. J. Cone, C. G. Sinon, S. M. Fortin, P. A. Kantak, I. B. Witten, K. Deisseroth, G. D. Stuber, M. F. Roitman, Optical suppression of drug-evoked phasic dopamine release. *Front. Neural Circuits* **8**, 114 (2014).
25. 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).
26. M. Mahn, I. Saraf-Sinik, P. Patil, M. Pulin, E. Bitton, N. Karalis, F. Bruentgens, S. Palgi, A. Gat, J. Dine, J. Wietek, I. Davidi, R. Levy, A. Litvin, F. Zhou, K. Sauter, P. Soba, D. Schmitz, A. Lüthi, B. R. Rost, J. S. Wiegert, O. Yizhar, Efficient optogenetic silencing of neurotransmitter release with a mosquito rhodopsin. *Neuron* **109**, 1621–1635.e8 (2021).

27. A. Mathis, P. Mamidanna, K. M. Cury, T. Abe, V. N. Murthy, M. W. Mathis, M. Bethge, DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci.* **21**, 1281–1289 (2018).
28. M. G. Kutlu, J. E. Zachry, P. R. Melugin, J. Tat, S. Cajigas, A. U. Isiktas, D. D. Patel, C. A. Siciliano, G. Schoenbaum, M. J. Sharpe, E. S. Calipari, Dopamine signaling in the nucleus accumbens core mediates latent inhibition. *Nat. Neurosci.* **25**, 1071–1081 (2022).
29. Y. K. Takahashi, H. M. Batchelor, B. Liu, A. Khanna, M. Morales, G. Schoenbaum, Dopamine neurons respond to errors in the prediction of sensory features of expected rewards. *Neuron* **95**, 1395–1405.e3 (2017).
30. M. P. H. Gardner, G. Schoenbaum, S. J. Gershman, Rethinking dopamine as generalized prediction error. *Proc. R. Soc. B. Biol. Sci.* **285**, 20181645 (2018).
31. N. Eshel, M. Bukwich, V. Rao, V. Hemmelder, J. Tian, N. Uchida, Arithmetic and local circuitry underlying dopamine prediction errors. *Nature* **525**, 243–246 (2015).
32. R. S. Lee, B. Engelhard, I. B. Witten, N. D. Daw, A vector reward prediction error model explains dopaminergic heterogeneity. bioRxiv 2022.02.28.482379 (2023). https://doi.org/10.1101/2022.02.28.482379.
33. Y. K. Takahashi, T. A. Stalnaker, L. E. Mueller, S. K. Harootonian, A. J. Langdon, G. Schoenbaum, Dopaminergic prediction errors in the ventral tegmental area reflect a multithreaded predictive model. *Nat. Neurosci.* **26**, 830–839 (2023).
34. D. A. Burke, H. Jeong, B. Wu, S. A. Lee, J. R. Floeder, V. M. K. Namboodiri, Few-shot learning: Temporal scaling in behavioral and dopaminergic learning. bioRxiv 2023.03.31.535173 (2023). https://doi.org/10.1101/2023.03.31.535173.
35. F. Carter, M.-P. Cossette, I. Trujillo-Pisanty, V. Pallikaras, Y.-A. Breton, K. Conover, J. Caplan, P. Solis, J. Voisard, A. Yaksich, P. Shizgal, Does phasic dopamine release cause policy updates? *Eur. J. Neurosci.* **59**, 1260–1277 (2023).
36. L. T. Coddington, S. E. Lindo, J. T. Dudman, Mesolimbic dopamine adapts the rate of learning from action. *Nature* **614**, 294–302 (2023).
37. 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).
38. M. G. Kutlu, J. E. Zachry, P. R. Melugin, S. A. Cajigas, M. F. Chevee, S. J. Kelley, 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).
39. R. S. Rodger, Multiple contrasts, factors, error rate and power. *Br. J. Math. Stat. Psychol.* **27**, 179–198 (1974).
40. P. Jean-Richard-dit-Bressel, C. W. G. Clifford, G. P. McNally, Analyzing event-related transients: Confidence intervals, permutation tests, and consecutive thresholds. *Front. Mol. Neurosci.* **13**, 14 (2020).
41. J. W. de Jong, Y. Liang, J. P. H. Verharen, K. M. Fraser, S. Lammel, State and rate-of-change encoding in parallel mesoaccumbal dopamine pathways. *Nat. Neurosci.* **27**, 309–318 (2024).
42. K. M. Fraser, H. J. Pribut, P. H. Janak, R. Keiflin, From prediction to action: Dissociable roles of ventral tegmental area and substantia nigra dopamine neurons in instrumental reinforcement. *J. Neurosci.* **43**, 3895–3908 (2023).

Acknowledgments: We thank members of the Janak laboratory including C. Drieu for helpful discussions on photometry analysis, A. Dong for help with video scoring and constructing optic fibers, and C. Shuai for help with running a subset of rats. We also thank M. Burrell and N. Uchida for the helpful discussion. We acknowledge G. Costa and A. Asiminas for making rat vector illustrations available on SciDraw (scidraw.io). **Funding:** This work was supported by NIH grants F32 DA054767 (to E.G.) and R01 DA035943 (to P.H.J.). **Author contributions:** E.G. and P.H.J. designed the experiments. E.G., Y.C., S.B., L.C., and R.M. collected the data. E.G. and P.H.J. visualized and analyzed the data with input from Y.C., H.J., and A.B. E.G. prepared the manuscript with input from P.H.J., V.M.K.N., H.J., and Y.C. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Data are available for download via Dryad: https://doi.org/10.5061/dryad.q573n5tr1.

Submitted 8 December 2023
Accepted 23 April 2024
Published 29 May 2024
10.1126/sciadv.adn4203