Few-shot learning: temporal scaling in behavioral and dopaminergic learning

Dennis A Burke[1], Huijeong Jeong[1], Brenda Wu[1], Seul Ah Lee[1, 2], Joseph R Floeder[3], Vijay Mohan K Namboodiri[1, 3, 4#]

[1] Department of Neurology, University of California, San Francisco, CA, USA

[2] University of California, Berkeley, CA, USA

[3] Neuroscience Graduate Program, University of California, San Francisco, CA, USA

[4] Weill Institute for Neurosciences, Kavli Institute for Fundamental Neuroscience, Center for Integrative Neuroscience, University of California, San Francisco, CA, USA

[#] Correspondence to VijayMohan.KNamboodiri@ucsf.edu

Abstract

How do we learn associations in the world (e.g., between cues and rewards)? Cue-reward associative learning is controlled in the brain by mesolimbic dopamine[1–4]. It is widely believed that dopamine drives such learning by conveying a reward prediction error (RPE) in accordance with temporal difference reinforcement learning (TDRL) algorithms[5]. TDRL implementations are "trial-based": learning progresses sequentially across individual cue-outcome experiences. Accordingly, a foundational assumption—often considered a mere truism—is that the more cue-reward pairings one experiences, the more one learns this association. Here, we disprove this assumption, thereby falsifying a foundational principle of trial-based learning algorithms. Specifically, when a group of head-fixed mice received ten times fewer experiences over the same total time as another, a single experience produced as much learning as ten experiences in the other group. This quantitative scaling also holds for mesolimbic dopaminergic learning, with the increase in learning rate being so high that the group with fewer experiences exhibits dopaminergic learning in as few as four cue-reward experiences and behavioral learning in nine. An algorithm implementing reward-triggered retrospective learning explains these findings. The temporal scaling and few-shot learning observed here fundamentally changes our understanding of the neural algorithms of associative learning.

Introduction

The neurobiological study of reward learning is dominated by the hypothesis that dopamine signals a temporal difference RPE[5] and that the brain implements TDRL[6,7]. TDRL has been hugely influential in the study of cue-reward learning, explaining numerous behavioral phenomena as well as dopamine dynamics across learning[1,8–16]. A major advance of TDRL over previous formal descriptions of reward learning such as the Rescorla-Wagner model[17] is the ability to account for the passage of time during the experience of cues and rewards. However, common TDRL formulations of animal learning only model time during a "trial period", an experimenter defined period from a cue through an outcome (e.g., reward delivery or omission), and do not consider the inter-trial interval (ITI)[5–7,15]. Thus, they implicitly assume that learning occurs in trials, and that the rate of learning is determined only by the trial period.

However, many fields have consistently noted that learning becomes more effective when experiences are more temporally spaced. This concept is so widely known that students are regularly advised that study sessions spread over time are more effective than "cramming" before an exam. Such spacing effects have been demonstrated across many domains of learning in species ranging from humans through invertebrates[18–26], including mammalian cue-reward learning[27–33]. In addition to these qualitative observations, it has even been suggested that associative learning is timescale invariant, in which the number of experiences required for learning is a function of the ratio between ITI and the cue-reward delay (ITI/trial ratio)[34–36]. However, most demonstrations of the effectiveness of spaced versus

massed cue-reward learning examine massed learning with a short ITI relative to the cue-reward delay[27,32] (ITI/trial ratio less than 10). Demonstrations like these do not rule out TDRL because under these conditions, extensions of TDRL that account for the ITI[37] can produce faster learning with longer ITIs (Extended Data Fig 1, 60 s vs. 6 s). Thus, to rigorously examine the predictions of the dominant neurobiological "trial-based" models, experiments testing categorical, falsifiable predictions shared by TDRL implementations should be identified.

Here, we show using simulations that when the ITI is sufficiently long relative to the cue-reward interval, a TDRL implementation accounting for ITI[37] predicts virtually no additional gain in learning rate (Extended Data Fig 1B, C). Indeed, when the ITI/trial ratio was compared between 48 and 480, there was virtually no improvement in the rate of learning for the more spaced condition. Thus, a falsifiable prediction of this TDRL model is that learning should not be affected by increasing ITIs when ITI/trial ratios are already as large as 48. However, most prior studies of trial spacing with long ITIs examined ITI/trial ratios less than 48[27–30,32], thereby necessitating further empirical tests.

Temporal scaling in behavioral learning

To directly test categorical, falsifiable predictions of these implementations of TDRL, we tested whether increases in ITI affect learning beyond large ITI/trial ratios. To this end, we classically conditioned thirsty head-fixed mice with similar parameters as the two longer ITI simulations described above. Mice were conditioned to associate a brief auditory tone (0.25 s, 12 kHz) with the delivery of sucrose solution reward (15% w/v, 2-3 μL) through a spout positioned in front of their mouth (Fig 1A). Two groups of mice were presented with this same trial structure, with one group, Typical ITI mice (Typ ITI), experiencing 60 s ITIs (ITI/trial ratio = 48) and another group, Extended ITI mice (Ext ITI), experiencing 600 s ITIs (ITI/trial ratio = 480). Both groups were trained for ~1hr per day. So, Typ ITI were presented 50 cue-reward pairings a day, while Ext ITI mice were presented 6 cue-reward pairings a day (this accounts for a fixed reward consumption period; see Methods). Both groups of mice were conditioned for at least 8 days. The head-fixed preparation is critical to test ITI/trial ratios of 48 and 480. By directing the mouse's attention to the spout, brief cues can be used, which allows for conditioning with very short trial periods relative to ITI. This approach also enables conditioning to begin without the need for pretraining mice to collect rewards, which can lead to the formation of other learned associations. Furthermore, head-fixation ensures equivalent experiences of cue and reward delivery since animals with different ITIs are equally positioned relative to reward spout.

Using these groups of mice, we tested between two hypotheses (Fig 1B). Hypothesis 1 is based on TDRL, which predicts that once the ITI is sufficiently longer than the trial, trial-by-trial learning should be equivalent (Extended Data Fig 1C). Because Typ ITI mice will experience 10 times more cue-reward pairings, they will show greater evidence of learning at the end of conditioning than Ext ITI mice. Hypothesis 2 is that prior suggestions of faster learning in "spaced learning"[28,29,32] (but see[38]) apply even when the ITI is much longer than trial duration (ratio of 48 vs 480). Here, we present a strict version of this hypothesis in which the group that experiences 10 times fewer trials learns 10 times more per trial. Stated differently, Hypothesis 2 is that deleting 9 out 10 experiences for Typ ITI mice, and thereby extending the ITI 10 times, has no influence on overall learning.

We measured behavioral learning using cue-evoked anticipatory licks before reward delivery[39–41]. Mice from both groups began to show cue-evoked licks in the first few days of conditioning (Fig 1C, Extended Data Fig 2A). When looking at cue-evoked licking as a function of cue-reward experiences, however, Ext ITI learned and reached asymptotic behavior in many fewer trials than Typ ITI mice (Fig 1D). By trial 40, Ext ITI mice showed significantly more cue-evoked licking (Typ ITI: 1.1 ± 0.4 Hz, Ext ITI: 3.7 ± 0.3 Hz, <0.0001; Fig 1D, Extended Data Fig 2B) and were significantly more likely to respond to the cue (Typ ITI: 0.29 ± 0.06, Ext ITI: 0.92 ± 0.04, <0.0001; Extended Data Figs 2C, 2D) than Typ ITI mice. This behavior is consistent with Hypothesis 2 that lengthening the time between cue-reward experiences improves learning even in conditions when the ITI is orders of magnitude longer than the trial duration (Fig 1B).

To fully compare learning rates between groups, we determined the first trial at which each individual showed evidence of learning using the cumulative sum of cue-evoked licks[3,30,42–45] (see Methods; Fig 1E, Extended Data Fig. 2). Remarkably, Ext ITI mice learned in ~9 trials on average (8.8 ± 0.6), significantly less than the 94 (94 ± 7) trials needed for Typ ITI mice to learn ($p<0.0001$; Fig 1F). By lengthening the ITI by a factor of 10, cue-reward learning required 10 times fewer trials, showing a quantitative scalar relationship between ITI duration and per-trial learning. This scalar relationship was not just limited to the learned trial number, as a single trial for Ext ITI mice was worth 10 trials for Typ ITI mice throughout the learning process (Figs 1G, 1H, Extended Data Fig 4A, B). Because Ext ITI mice have the same experience as Typ ITI but with the deletion of 9 out of 10 trials (i.e., 10 times the ITI), the overlap of the learning curves demonstrates that those "deleted" trials have no effect on learning.

Further suggesting that learning between groups was simply scaled, average asymptotic cue-evoked lick rates (Typ ITI: 4.06 ± 0.54 Hz, Ext ITI: 3.80 ± 0.26 Hz, $p = 0.66$; Fig 1I), the likelihood of responses to the cue (Typ ITI: 0.77 ± 0.08, Ext ITI: 0.92 ± 0.03, $p = 0.098$; Extended Data Fig 4C), and the abruptness of change, a measure of the steepness of individual animal learning curves (Typ ITI: 0.18 ± 0.02, Ext ITI: 0.18 ± 0.02, $p = 0.97$; Extended Data Fig 4D), were all similar between groups at the end of conditioning. Interestingly, despite similar average rates of asymptotic cue-evoked licking, Typ ITI mice showed significantly more variance in individual behavior compared to Ext ITI at the end of conditioning ($p<0.01$; Figs 1H, 1I; two Typ ITI mice did not learn the cue-reward association, Extended Data Fig 3C). This variance was also seen in the number of trials to learn when comparing mice that did show evidence of learning ($p<0.0001$; Fig 1F). This shows that individual variability in learning is driven in part by the environment and is not just a reflection of innate abilities.

This scalar relationship between ITI duration and learning is categorically inconsistent with trial-based accounts of learning. However, Ext ITI mice differ from Typ ITI mice in both duration of ITI and in number of trial experiences a day. This could lead Ext ITI mice to experience cues and rewards as more salient due to their sparsity and correspondingly higher level of novelty, despite being identical to those experienced by Typ ITI mice. As more salient stimuli can lead to greater conditioning in trial-based accounts of learning[46,47], this is a possible way in which trial-based accounts of learning could explain our results. To test this hypothesis, we conditioned a third group of mice (Typ ITI-few) with the same ITI as Typ ITI mice (mean: 60 s) and the same number of trials per day as Ext ITI mice (six). Across trials, learning in these mice progressed similarly to Typ ITI mice (Extended Data Fig 5). Late in conditioning, Typ ITI-few licked significantly less to the cue than Ext ITI mice (Typ ITI-few: 0.7 ± 0.3 Hz, Ext ITI: 3.7 ± 0.3 Hz, $p<0.0001$; Extended Data Fig 5B), similar to Typ ITI mice during the same trial numbers (Typ ITI-few: 0.7 ± 0.3 Hz, Typ ITI: 1.1 ± 0.4 Hz, $p = 0.38$; Extended Data Fig 5B). These results show that the difference in learning between Typ ITI and Ext ITI mice was not due to differences in cue novelty.

Temporal scaling in dopaminergic learning

The dominance of trial-based accounts of associative learning is supported in large part by the concordance between mesolimbic dopamine signaling and the error term in TDRL models. In temporal difference cue-reward learning, the goal is to estimate the value of a cue, which is used to drive behavior. By acting as an error signal for continuous updates to the value function, dopamine should therefore be tightly coupled to behavior[5,39,48]. Thus, to understand how our results of temporal scaling fit with current conceptions of associative learning, it is important to understand how dopamine signaling evolves over the course of learning in both Typ ITI and Ext ITI mice. Given the vastly different number of trials to acquisition in each group, we hypothesized two possible relationships between dopaminergic and behavioral learning (Fig 2B). Hypothesis 1 is that the development of cue-evoked dopamine (dopaminergic learning) precedes the emergence of behavior by a fixed number of trials in both groups. This hypothesis relies on the previously mentioned dopamine model, which proposes a strong connection between cue-triggered dopamine and behavioral learning. According to this model, the growth of dopaminergic cue responses reflects underlying increases in cue value, which drives behavioral learning. Because Ext ITI mice learn in ten times fewer experiences than Typ ITI mice (Fig 1F), Hypothesis

2 is that the development of cue-evoked dopamine also occurs in ten times fewer experiences and hence precedes behavioral learning by ten times fewer experiences.

To test these hypotheses, we measured dopamine release in the nucleus accumbens core in a subset of Typ ITI and Ext ITI animals with fiber photometry recordings of the optical dopamine sensor dLight1.3b (Fig 2A, Extended Data Fig 6). As can be seen in example mice, dopamine was evoked by reward receipt beginning on the first trial, but cue-evoked dopamine release developed over trials and preceded the emergence of behavioral learning, in line with prior work[49–51] (Fig 2C, Extended Data Fig 7A). To determine the trial at which cue-evoked dopamine emerged, we applied the same algorithm used to determine the learned behavior trial on the cumulative sum of the cue-evoked dopamine (Fig 2D, Extended Data Fig 7B; see Methods). Again, we found the same quantitative scalar relationship between ITI duration and dopaminergic learning. Dopaminergic learning in Ext ITI mice began on average between trials three and four (3.6 ± 0.4), significantly earlier than Typ ITI mice, which began to show cue-evoked dopamine responses at trial 36 (36 ± 7) ($p<0.05$; Extended Data Fig 7C). We then calculated the lag between dopaminergic and behavioral learning by subtracting the dopamine learned trial from the behavior learned trial in each individual mouse. In Ext ITI mice, dopaminergic learning precedes behavior by 5 trials on average (5.0 ± 0.7), significantly fewer than the 59 (59 ± 7) trials between dopaminergic and behavioral learning in Typ ITI mice ($p<0.01$; Fig 2E). Thus, per trial development of cue-evoked dopamine responses also scales with the duration of the ITI in learning: by increasing the ITI by a factor of ten, cue-evoked dopamine appears in ten times fewer trials and precedes behavioral learning in ten times fewer trials (Fig 2F, Extended Data Fig 8A). This scaling is consistent with Hypothesis 2 (Fig 2B).

Interestingly, despite the scaling in the onset of dopaminergic learning, cue-evoked dopamine in Ext ITI mice rose to asymptotic levels more rapidly and increased by more than a factor of ten per trial as compared to Typ ITI mice (Fig 2G, Extended Data Figs 8B, 8E, 8F). Asymptotic cue-evoked dopamine (relative to maximum reward response) was also significantly higher at the end of conditioning in Ext ITI mice compared to Typ ITI mice (0.47 ± 0.06 vs. 0.31 ± 0.02, $p<0.05$; Fig 2H). Furthermore, although there were differences in dopamine dynamics and learning rates between the groups, we observed a similar pattern in the dopamine reward response in both groups. Specifically, the reward response did not start at its maximum value during the first trial but rather increased during early conditioning, reaching its peak prior to the onset of behavior (Extended Data Figs 8C, 8D). This increase in reward-evoked dopamine across early experiences with reward has been noted before[3,52], and is further inconsistent with TDRL models of dopamine function. In the TDRL framework, the first experience of a particular reward, as is the case for trial 1 in our experiments, should evoke the maximum dopamine response across conditioning due to its completely unpredicted occurrence.

A model of retrospective causal learning explains temporal scaling

Because our TDRL simulations proved inadequate in predicting greater learning per experience even when the entire ITI was modeled (Extended Data Fig 1), we sought a different framework for explaining our results. We recently proposed a new formal model of dopamine-driven associative learning where associations are formed by retrospectively inferring the cause of rewards[3]. In this model, animals learn cue-reward associations through calculation of an adjusted net contingency for causal relations (ANCCR) based on estimates of the rate of cues at reward times vs. the baseline rate of cues. Cue presentations evoke an exponentially decaying eligibility trace. After receiving reward, the “memory” of these cues, represented by their eligibility traces, is used to estimate both the rate of cues at time of reward delivery and the overall rate of cues in the environment. These allow the animal to determine whether reward is contingent on cues. Within this framework, one possible explanation for the different rates of dopaminergic and behavioral learning seen in Typ ITI and Ext ITI mice is that the eligibility trace (memory) of the cue has fully decayed before the next cue-reward pairing in Ext ITI conditioning but is still active on subsequent trials for Typ ITI mice. If so, learning will be slower for Typ ITI mice. Cue rate

at reward time and baseline will both be high in Typ ITI, slowing down contingency estimates relative to Ext ITI mice where estimates of the cue rate at baseline will be near zero (Fig 3A).

To test this hypothesis, we ran ANCCR simulations using the exact trial (1.25 s) and ITI durations that were used to condition Typ ITI (60 s ITI) and Ext ITI (600 s ITI) mice—the same durations which TDRL simulations predicted would lead to learning after a nearly equal number of trials (Typ ITI: 92.0 ± 0.1, Ext ITI: 91.5 ± 0.1, $p<0.01$; Fig 3B, Extended Data Fig 1C). Here, we set the decay constant for the eligibility trace to 200 seconds (see Methods for details), allowing previous cue-reward pairings to influence rate calculations for a 60 s ITI, but not for a 600 s ITI. Remarkably, ANCCR simulations with a Typ ITI took 126 (126 ± 1) trials to learn, while simulations with an Ext ITI took ~12 trials (11.7 ± 0.2), capturing the experimentally observed temporal scaling between ITI duration and learning rate ($p<0.0001$; Fig 3C). This scaling was also found in dopaminergic learning (Typ ITI: 82 ± 0.9, Ext ITI: 7.5 ± 0.2, $p<0.0001$; Extended Data Fig 9A) and the lag between dopaminergic learning and the onset of behavior (Typ ITI: 44 ± 0.8, Ext ITI: 4.3 ± 0.2, $p<0.0001$; Fig 3D). Importantly, all parameters aside from the duration of the ITI were equivalent between groups. ANCCR also accurately predicted the greater cue-evoked dopamine at the end of conditioning seen in Ext ITI mice (Typ ITI: 0.56 ± 0.001, Ext ITI: 0.74 ± 0.006, $p<0.0001$; Fig 3E), but differed from experimental observations by also predicting slightly higher cue-evoked licking in Ext ITI mice (Typ ITI: 0.74 ± 0.001, Ext ITI: 0.82 ± 0.002, $p<0.0001$; Extended Data Fig 9B). Nevertheless, ANCCR captures the primary experimental findings, establishing a framework to comprehend temporal scaling of dopaminergic and behavioral cue-reward learning.

Discussion

We show that the rate of cue-reward learning scales quantitatively with the time between consecutive cue-reward experiences (Fig 1), requiring a reevaluation of the frameworks used to describe associative learning. These results are categorically inconsistent with “trial-based” models of learning, and require a reassessment of the implicit assumption that the trial is the fundamental unit of learning[32]. However, these data are consistent with prior work, primarily based on pigeon autoshaping paradigms, that has suggested that cue-reward learning is timescale invariant[34,35]. Importantly, our work extends this by demonstrating that the scaling applies even over very large ITI/trial ratios never tested in cue-reward conditioning (48 vs. 480) and that this effect obeys a quantitative scaling law. Increasing the ratio by a factor of ten leads to ten times more learning per trial. Any formal description of learning must explain not just greater learning with longer ITI, but also the quantitative scaling observed here. We show that ANCCR, a retrospective causal learning model, explains this quantitative scaling (Fig 3).

Due to this quantitative scaling, we demonstrate that auditory cue-reward conditioning can occur in few experiences. While some tasks such as fear conditioning[53], Morris water maze[54], or mate detection[55] are learned in few experiences, auditory cue-reward conditioning typically takes hundreds of trials[41,56]. While it is widely believed that fast learning in some of these tasks is due to the salience of outcomes such as shocks, drowning, and mating, the current results raise the intriguing possibility that this difference may reflect the low frequency of these outcomes in the lives of animals. It remains to be tested whether such outcome frequency effect might explain the wide variety of learning rates in naturalistic and laboratory tasks[57,58].

While TDRL implementations that ignore the ITI evidently cannot explain ITI effects, we demonstrated that an extension of TDRL that includes the ITI is also unable to account for the experimentally observed temporal scaling. A goal of TDRL simulations in neuroscience is to fit the fast timescale fluctuations of dopamine. Due to the timescale of phasic dopamine dynamics (hundreds of milliseconds) and the brief trial period (1.25 s from cue onset to reward), the time resolution needed to accurately account for intra-trial dopamine dynamics requires hundreds of states at minimum to model the ITI even in Typ ITI mice, the “shorter” ITI tested here. Thus, it is unlikely that ITI states could acquire differential value to drive the quantitative differences in learning observed here. Therefore, though we cannot rule out future TDRL extensions, any model in which time is represented through a series of states is unlikely to explain our data.

One possibility by which “trial-based” frameworks of learning could be used to account for the data presented here is to assume that replays/reactivations of cue-reward experiences during the extended ITI provides “virtual trials” in lieu of the real trials experienced by mice[56,59]. While we cannot rule out this possibility, the similarity between the learning curves for the long and short ITI groups suggests that such a mechanism would somehow have to precisely replicate the effect of the missing 9/10 experiences in the long ITI group, and the available evidence suggests that neither the structure of replay/reactivation events nor the information they encode perfectly replicate past experiences[60–66]. We therefore suggest that our proposed non-trial-based learning rule is a more parsimonious explanation of the *quantitative* scaling observed here.

A potential alternate qualitative explanation for our results might be that the stimulus salience is considerably higher in the Ext ITI group compared to Typ ITI, either due to increased novelty of the cue (resulting from fewer total experiences), or lower habituation across repeated presentations. However, neither explanation is consistent with this or prior studies. Specifically, novelty induced salience is ruled out by the fact that the extended ITI group learns ten times faster than the Typ ITI-few group (Extended Data Fig 5) despite receiving equal number of cue presentations (which equates cue novelty between groups). Similarly, habituation of cue salience is ruled out because prior studies have shown that cue-induced pupil responses are stable and do not habituate with repeated trials even with ITIs as short as 5-20 s[67–69]. Further, there is little dopaminergic habituation beyond 60 s[70]. Nevertheless, it is worth mentioning that a verbal description of our explanation (Fig 3) could be that cues are more “salient” in the Ext ITI group (i.e., have lower baseline rates and hence a higher “temporal salience”).

In addition, we establish that dopaminergic learning exhibits temporal scaling, whereas the asymptotic response does not (Fig 2), thereby offering a significant new constraint for dopamine-mediated learning models. While prior work has provided accumulating evidence that mesolimbic dopamine signals do not function strictly as a TDRL reward prediction error signal[3,52,71–76], the current results call into question the broader trial-based reinforcement learning framework used to understand dopamine and learning. While some prior models do explain the quantitative scaling of behavioral learning, these models do not yet explain dopaminergic dynamics[35,36,77,78]. Collectively, we provide a new framework for understanding dopamine mediated cue-reward learning[3] that explains temporal scaling in both dopaminergic and behavioral learning (Fig 3). As learning updates occur at every reward, our model does not rely on the concept of an experimenter defined trial, which leads to many problematic assumptions[79]. Thus, it is uniquely suited to account for experience outside an arbitrarily imposed “trial period”. In this way, the theory gets us closer to explaining naturalistic learning outside laboratories where experiences do not have defined trial structures. By providing a framework for understanding the data presented here, grounded in the known dynamics of dopamine mediated learning[3], our results provide further support for a reevaluation of the neural algorithms underlying learning.

Acknowledgments

We thank J. Berke, L. Frank, M. Kheirbek, G.D. Stuber, M. Andermann, S. Mihalas, I. Trujillo Pisanty, J. Rodriguez-Romaguera, and members of the Namboodiri laboratory for helpful discussions. This project was supported by NIH R00MH118422, R01MH129582, R01AA029661, and the Scott Alan Myers Endowed Professorship (V.M.K.N.). The authors have no competing interests.

Author contributions

D.A.B. and V.M.K.N. conceived the project. D.A.B., H.J., B.W., S.L., and J.R.F. performed experiments. D.A.B. performed analyses. H.J. performed simulations. V.M.K.N. oversaw all aspects of the study. D.A.B. and V.M.K.N. wrote the manuscript with help from all authors.

Methods

Animals

All experiments and procedures were performed in accordance with guidelines from the National Institutes of Health Guide for the Care and Use of Laboratory Animals and approved by the UCSF Institutional Animal Care and Use Committee. Fifty adult (>11 weeks at time of experiments; median: 13 weeks) wild-type male and female C57BL/6J mice (JAX; RRID:IMSR_JAX:000664) were used across three experimental groups: Extended ITI (n = 19; 12 behavior-only [5F/7M] and 7 DA+behavior [5F/2M]), Typical ITI (n = 19; 14 behavior-only [8F/6M] and 5 DA+behavior [3F/2M]), and Typical ITI – few trials (n = 12; 6F/6M). Two Typical ITI mice were implanted with optic fibers, but excluded from dopamine analysis for either failing to learn the cue-reward association (Mouse 46, Extended Data Fig 3; see below) or due to a missed fiber placement (Mouse 50, Extended Data Fig 6).

All mice were head-fixed during conditioning and underwent surgery prior to behavior experiments to either implant a custom head-ring for head-fixation (behavior-only) or to inject viral vector and implant an optic fiber and head-ring (DA+behavior) (See Surgery section). Mice were at minimum 8 weeks old at time of surgery (median: ~9.5 weeks). Following surgery, mice were given at least a week to recover before beginning water deprivation. During water deprivation, mice were given ad libitum access to food but were water deprived to ~85 – 90% of pre-deprivation bodyweight and maintained in that weight range throughout experiments through daily adjustments to water allotment. Mice were weighed and monitored daily for the duration of deprivation.

After surgery, mice with only a head ring implant were group housed in cages containing mice from multiple experimental groups, while fiber implanted mice were single housed. Mice were housed on a reverse 12-h light/dark cycle, and all behavior was run during the dark cycle.

Surgery

Surgery was performed under aseptic conditions. Mice were anesthetized with isoflurane (5% induction, ~1-2% throughout surgery) and placed in the stereotaxic device (Kopf Instruments) and kept warm with a heating pad. Prior to incision, mice were administered carprofen (5 mg/kg, SC) for pain relief, saline (0.3 mL, SC) to prevent dehydration, and local lidocaine (1 mg/kg, SC) to the scalp for local anesthesia. All mice were implanted with a custom-designed head ring (5 mm ID, 11 mm OD, 3 mm height) on the skull for head-fixation. The ring was secured to the skull with dental acrylic supported by screws. Following surgery, mice were given buprenorphine (0.1 mg/kg, SC) for pain relief.

To measure dopamine release in a subset of mice, 500 nL of an adeno-associated viral (AAV) vector encoding the dopamine sensor dLight1.3b (AAVDJ-CAG-dLight1.3b, 2.4 x 10^{13} GC/ml diluted in sterile saline to final titer of 2.4 x 10^{12} GC/ml) was injected unilaterally into NAc core (from bregma: AP 1.3, ML +/-1.4, DV -4.55), in either right or left hemisphere, counterbalanced across groups. Viral vectors were injected through a small glass pipette with a Nanoject III (Drummond Scientific) at a rate of 1 nL/s. Injection pipette was kept in place 5-10 min to allow diffusion, then slowly retracted to prevent backflow up the injection tract. Following injection, an optic fiber (NA 0.66, 400µm, Doric Lenses) was implanted 200-350 µm above the virus injection site. Following fiber implant, the head ring was secured to skull as above. Following the conclusion of conditioning, fiber implanted mice were transcardially perfused, and brains were fixed in 4% paraformaldehyde. Brains were sectioned at 50 µm and imaged on a Keyence microscope to verify fiber placement.

Conditioning

All animals were conditioned with an identical trial structure, differing only in inter-trial interval (ITI) and/or number of trial presentations. Typical ITI (Typ ITI) mice were run for 50 trials a day with a variable ITI with a mean of 60 s (uniformly distributed from 48 s to 72 s). Extended ITI (Ext ITI) mice were run for 6 trials a day with a variable ITI with a mean of 600 s (uniformly distributed from 480 s to 720 s).

Because Ext ITI mice experienced fewer trials a day to keep total conditioning time roughly equal between groups, a third group, Typical ITI – few trials (Typ ITI-few), was run for 6 trials a day with a mean of 60 s (uniformly distributed from 48 s to 72 s; same as Typ ITI) to control for the difference in total trial experiences between Typ ITI and Ext ITI mice.

Trials consisted of a 0.25 s 12 kHz constant tone through a piezo speaker followed by a 1 s delay (trace period) after which sucrose sweetened water (2- 3 µL; 15% w/v) was delivered through a gravity fed solenoid to a lick spout in front of the mouse, controlled by custom Matlab and Arduino scripts. After each trial, there was a fixed three second period following reward delivery to allow reward consumption. Though this is technically a part of the ITI, we omitted this interval when calculating ITI/trial ratios for simplicity. Lick spout was positioned close to the animals such that animals could sense, but were not touched by, delivery of reward. Licks were detected through a complete-the-circuit design and recorded in Matlab.

Mice were not habituated to the head-fixation apparatus or sucrose delivery prior to conditioning. For the vast majority of mice, the first trial was their first experience of liquid sucrose reward. An initial subset of behavior only Ext ITI mice (n = 6) ran with a fixed ITI of 600 s and was given a single uncued reward delivery prior to conditioning on day 1. No gross difference in learning compared to subsequent groups was detected, and data were pooled. For all other groups on day 1, mice were placed in the head-fixation apparatus and conditioning commenced. Because a minority of animals from each condition did not initially consume sucrose at time of reward delivery, for all analysis, “trial 1” was defined as the first trial in which a mouse licked to consume sucrose within 5 seconds of reward delivery. Mice were run for at least 8 days of conditioning, and trial analyses included the first 40 trials (Ext ITI) or 400 trials (Typ ITI).

Fiber photometry

To measure dLight signal, light from 470 nm and 405 nm LEDs integrated into a fluorescence filter minicube (Doric Lenses) was passed through a low-autofluorescence patchcord (400 µm, 0.57 NA, Doric Lenses) to the mouse. Emission light was collected through the same patchcord, bandpass filtered through the minicube, and measured with a single integrated detector. Excitation LED output was sinusoidally modulated by a Doric Fiber Photometry Console running Neuroscience Studio v5.4 at 530 hz (470) and 209 hz (405). The console demodulated the incoming detector signal producing separate emission signals for 470 nm excitation (dopamine) and 405 nm excitation (dopamine-insensitive isosbestic control). Signals were sampled at 12 kHz and subsequently downsampled to 120 Hz following low-pass filtering at 12 Hz. Due to a software error during photometry data file save, the final trial was not recorded on two occasions (1 Typ ITI, 1 Ext ITI) and was excluded from analysis. This error occurred either well before (Typ ITI) or well after (Ext ITI) the emergence of learning, and thus had minimal effect on the resulting analysis. A TTL pulse signaling behavior session start and stop was recorded by the photometry software to sync and align photometry and behavior data recorded on different hardware.

Analysis

<u>Behavior:</u> The behavioral measure of learning here was licking in response to the cue before reward delivery. As mice learn the cue-reward association, cue presentation elicits anticipatory licking behavior toward the reward spout. To measure the cue-evoked change in licking behavior over baseline, the number of licks in the 1.25 s baseline period before cue onset was subtracted from the number of licks in the 1.25 s period from cue onset to reward delivery to calculate the change in licking behavior to the cue (cue-evoked licks). When this number was converted to a rate, it was reported as “Δ lick rate to cue.” To binarize cue-evoked licking behavior, we also measured the proportion of mice in each group that made more than one cue-evoked lick on each trial across conditioning (Extended Data Figs 2 & 4). To visualize average trial licking behavior for each session in example animal plots (Figs 1C & 2C, Extended Data Figs 2A & 7B), lick peri-stimulus time histograms were generated by binning licks into

0.1 s bins, converting to a rate, and averaging across trials. The resulting average lick rate trace was smoothed with a Gaussian filter to aid visualization.

To calculate the trial at which animals show evidence of learning, we first took the cumulative sum (cumsum) of the cue-evoked licks[3,30,42–44]. Then drawing a diagonal from beginning to the end of the cumsum curve, we calculated the first trial that occurred within 75% of the maximum distance from the curve to the diagonal, which corresponded to the trial at which cue-evoked licking behavior emerged (Extended Data Figs 3A-C). This trial was designated the "learned trial." Occasionally after learning, the number of a mouse's licks to cue tapers off. If at the calculated learned trial the diagonal line was underneath the cumsum curve, which means that the mouse's lick behavior was decreasing at that point rather than increasing, we iteratively reran the algorithm by drawing the diagonal from the beginning to the point on the cumsum curve corresponding to the previously calculated trial until at the new calculated trial the diagonal was above the cumsum curve (corresponding to the trial in which lick behavior begins to increase). Note that we use the first trial within 75% of maximum distance rather than the overall maximum distance (which would be the largest inflection point in the curve) to account for variability in post-learning behavior that occasionally caused the maximum distance from the diagonal to be at a point after a mouse has consistently licked to the cue for many trials; however, this choice did not affect the main conclusion of the analysis that Ext ITI mice learn in ten times fewer trials than Typ ITI mice (Extended Data Fig 3D). Mice that did not show a > 0.5 Hz average increase in lick rate to cue for at least 2 sessions were classified as nonlearners and were not considered in comparison of learned trials (Fig 1F, Extended Data Figs 3C, 3D). To measure the steepness of individual animal learning curves, we calculated the abruptness of change at the learned trial as the distance from the cumsum curve to the diagonal described above. This distance was calculated in normalized units where the top of the diagonal was set to equal 1. Cumsum data is occasionally displayed normalized to the number of trials (yielding a y-axis that corresponds to average response across all prior conditioning trials) to better compare across groups that experienced different numbers of trials.

<u>Dopamine:</u> To analyze the signals, a session-wide dF/F was calculated by applying a least-squares linear fit to the 405 nm signal to scale and align it to the 470 nm signal. The resulting fitted 405 nm signal was then used to normalize the 470 nm signal. Thus, dF/F is defined as dF/F = (470 nm signal − fitted 405 nm signal)/fitted 405 nm, expressed as a percent [80]. Cue-evoked dopamine (cue DA) was measured as the area under the curve (AUC) of the dopamine signal for 0.5 s following cue onset minus the AUC of the baseline period 0.5 s directly preceding cue onset. Reward evoked dopamine (reward DA) was measured as the AUC 0.5 s following the first detected lick after reward delivery minus the AUC of the pre-cue baseline period described above. If the onset and offset of a detected lick spanned reward delivery time, the reward AUC was calculated from time of reward delivery. To facilitate comparisons across mice with differing levels of virus expression, dopamine measurements per mouse were normalized to the average of the three maximum reward responses in that mouse. Maximum rather than initial reward responses were chosen, as the reward response initially increased across early conditioning trials with different numbers of trials until maximum between conditions (Extended Data Fig 8C, D). All dopamine responses reported in main figures are AUC measurements, but peak measurements are also plotted as a comparison point (Extended Data Fig 8). To measure cue and reward peak dopamine responses, the mean dopamine signal during the baseline period was subtracted from the maximum value of the dopamine signal during the cue and reward windows described above for AUC measurements. Similar to AUC measurements, peak responses were also normalized to the mean of the max 3 reward responses in each animal.

To calculate the trial at which dopamine responses to the cue develop (DA learned trial), we took the cumsum of the normalized cue DA response described. A diagonal was drawn from trial 1 through the point on the cumsum curve at 1.5 times the behavior learned trial to account for decreasing cue responses with extended training[81]. The same algorithm described above to determine the behavior learned trial was run on the cue DA curve. The lag between DA and behavioral learned trial (the number

of trials between the development of dopamine responses to the cue and the emergence of behavioral learning) was defined as the behavior learned trial minus the DA learned trial (Fig 2E).

For one Typ ITI dLight animal, during an initial conditioning session, a software crash caused the loss of lick data for 50 trials experienced by the animal. An additional 13 trials were presented to the animal that day and recorded following the crash. Photometry data were recorded for all 63 trials. Because the crash occurred prior to the emergence of learning and cue-evoked licking behavior (as confirmed by both online observation by experimenter prior to crash and a -0.14 Hz average cue-evoked change in lick rate for the 13 trials recorded after crash), the 50 trials in which data were lost were coded as 0 cue-evoked licks. All 63 trials the animal experienced were included in analyses.

To visualize the average relationship between DA responses and licking behavior across learning with variability in individual learning rates, signals were aligned to behavior learned trial and plotted through 250 or 25 trials after learning (Fig 2, Extended Data Fig 8). For aligned cumsum plots, data were normalized by the value from trial 400 (Typ ITI) or trial 40 (Ext ITI).

Simulations

Temporal difference reinforcement learning (TDRL) simulation: TDRL assumes that animals assign value to each moment following an event (e.g., cue) to predict future reward. Each event elicits multiple states, and the value of each time step can be expressed as a weighted sum of activated states at that moment. If the prediction from previous moment is different from what is experienced in the current moment, the model updates the value of previous moment based on this reward prediction error, assumed to be signaled by dopamine. Depending on how the model represents a state, TDRL can be further divided into subtypes. Here, we used the microstimulus model[37] as a representative of TDRL because it naturally accounts for ITI as a set of states are triggered following reward delivery. This model assumes that time states are Gaussian functions of increasing width following each event (cue or reward). The following model parameters were used: bin size = 0.1s, learning rate (α) = 0.01, temporal discounting factor (γ) = 0.999, decay parameter of eligibility trace (λ) = 0.95, number of states elicited by each event (m) = 20, width of Gaussian function (σ) = 0.08, and decay parameter of event memory (d) = 0.99.

For the simulation in Extended Data Fig 1, we used a similar set of task parameters as used in later behavior experiments: 1.25-s cue-reward delay, 3-s post-reward delay, and uniformly distributed intertrial intervals between ±20% from the mean. Three different means of ITI were tested (6s, 60s, and 600s). Each case was iterated 20 times.

Adjusted net contingency for causal relation (ANCCR) simulation: We previously proposed a new learning model called ANCCR based on the learning of retrospective associations[3]. ANCCR operates by identifies cues that cause meaningful events such as reward by looking back from reward. The meaningfulness of an event can be expressed as a sum of innate meaningfulness (e.g., is this event innately rewarding or punishing?) and learned meaningfulness (e.g., does this event cause another meaningful event?). In ANCCR, we previously assumed that dopamine signals the learned meaningfulness of an event, and that innate meaningfulness is conveyed by another system. If the summed meaningfulness of a given event crosses a certain threshold, animals will start searching for the cause(s) of that event. Such meaningful events were labeled as meaningful causal targets (MCTs). Once the cause of an MCT is revealed, the dopamine response to the MCT is adjusted as the presence of the MCT is now "explained" by its cause. This framework captured the nature of mesolimbic dopamine release in our previous paper and current experiments, but it could not explain one aspect of the results: relatively high dopamine response to the first experience of reward (though this response increases with repeated reward experience, consistent with our previous demonstration[3]). To address this, we updated our previous ANCCR model with a slight tweak that is explained below.

According to ANCCR, dopamine response to reward can be expressed as shown below until its cause is revealed:

$$DA_r = [wC_{\rightarrow rr} + (1-w)C_{\leftarrow rr}]R_r \quad (1)$$

where $C_{\leftarrow rr}$ is the retrospective r←r association, $C_{\rightarrow rr}$ is the prospective r→r association, w is a weight (between 0 and 1), and $R_r$is the reward magnitude. In ANCCR, the prospective association between two events i and j ($C_{\rightarrow ij}$) is derived from the retrospective association ($C_{\leftarrow ij}$) by multiplying the ratio between baseline rates of two events ($M_{\leftarrow j-}/M_{\leftarrow i-}$). Thus, we assumed that the model does not derive prospective association and instead keeps it as zero when the rate of event i ($M_{\leftarrow i-}$) is negligibly small, which was set to be below 10^{-4} in our simulation. This property makes dopamine response to the first reward as below:

$$DA_r = [w \times 0 + (1-w)C_{\leftarrow rr}]R_r = (1-w)C_{\leftarrow rr}R_r \quad (2)$$

This in turn can be expressed as below:

$$DA_r = (1-w)C_{\leftarrow rr}R_r = (1-w)(M_{\leftarrow rr} - M_{\leftarrow r-})R_r \quad (3)$$

where $M_{\leftarrow rr}$ is the average reward rate at rewards and $M_{\leftarrow r-}$ is baseline reward rate. The model updates $M_{\leftarrow rr}$ every time an animal experiences reward using:

$$M_{\leftarrow rr} \equiv M_{\leftarrow rr} + \alpha(E_{\leftarrow rr} - M_{\leftarrow rr}) \quad (4)$$

where ≡ denotes an update operation, α is learning rate, and $E_{\leftarrow rr}$ is the eligibility trace of reward. With no prior experience of reward, the prior estimate of $M_{\leftarrow rr}$ is zero and $E_{\leftarrow rr}$ is one, because it only accounts for the current reward. Thus, the new estimate of $M_{\leftarrow rr}$ at the first reward will be:

$$M_{\leftarrow rr} \equiv 0 + \alpha(1-0) = \alpha \quad (5)$$

Since the baseline reward rate ($M_{\leftarrow r-}$) is continuously updated (in our simulation, we do this once every 0.2 s) regardless of the actual reward schedule, $M_{\leftarrow r-}$ at the first reward delivery is zero, which is the value updated prior to the reward delivery and does not consider the current reward. Thus, equation (3) can be expressed as below:

$$DA_r = (1-w)(\alpha - 0)R_r = (1-w)\alpha R_r \quad (6)$$

Given that w is a weight between 0 and 1 and α is typically set to a small value (<0.3, often less than 0.1 at the asymptote of learning), dopamine response to the first reward should be small (at most 0.2 times of reward magnitude, when w=0 and α=0.2). However, we observed a significantly positive dopamine response to the first reward, which was about ~50-60% of the largest reward response over learning (Extended Data Fig 8). To account for this mismatch, we updated our model by postulating that dopamine response partially accounts for the 'innate meaningfulness' (b) of a given stimulus in addition to its learned meaningfulness. Therefore, dopamine response to reward in the updated model is:

$$DA_r = [wC_{\rightarrow rr} + (1-w)C_{\leftarrow rr}]R_r + b_r \quad (7)$$

More generally, dopamine response to any given stimulus is expressed as below, which is the updated version of equation (17) in our prior study[3]:

$$DA_i = \sum_j \hat{C}_{\leftrightarrow ij}\, \boldsymbol{I}(j \in MCT) + b_i \quad (8)$$

where $\hat{C}_{\leftrightarrow ij}$ denotes ANCCR between i and j and $\boldsymbol{I}(j \in MCT)$ denotes that j is a meaningful causal target. This postulate that the innate meaning of a stimulus/reward is conveyed in part by dopamine and in part by another system is consistent with a previous experimental demonstration that a dopaminergic and non-dopaminergic pathway collectively encode aspects of reward[82]. Other calculations in the model remained the same as shown in our previous paper. For simulations, the following parameters were used: w=0.5, b_{cue}=0, b_{reward}=0.5, $threshold$=0.4, T=200 s, k=0.005, α_R=0.2. When an animal experiences a new task for the first time, we assumed that α will be higher than that after the animal

gets used to the task, consistent with standard meta-learning assumptions. Thus, we set α to start from 0.3 and exponentially decay with exponential parameter 0.1 until it reaches 0.02.

To simulate the experiments using ANCCR, we used the same task parameters as used in actual experiments. To approximate animal behavior for Extended Data Fig 9A, the probability of lick to cue was calculated by applying a softmax function:

$$p(lick|cue) = \frac{e^{V_{cue}/t}}{e^{V_{cue}/t}+e^{V_{nolick}/t}} \quad (9)$$

where t is the temperature and V_{cue} is a value of cue. V_{cue} was defined as the product of the prospective association between cue and reward, and the estimated magnitude of reward ($C_{\rightarrow cr}R_{cr}$), subtracted by the cost of action. We set the value of null action (no lick) as zero and cost of action as 0.3. Temperature (t) was set as 0.2. Given the much smaller noise in simulation than experimental data, the learned trial was simply defined as the trial in which the cumulative sum of behavior response was farthest from the diagonal. Other than that, simulation data were analyzed in the same way as experimental data.

It is worth noting that ANCCR has multiple possible explanations for the data observed here. The above explanation is one in which model parameters were assumed to be identical between the two ITI groups. Another explanation can be derived from first principles by postulating that the overall rate of learning over absolute time should be equal for learning the baseline rate of cues and the reward-triggered rate of cues. However, our purpose here is to show that even with simple assumptions of model parameters that were identical in both groups, ANCCR produces quantitative temporal scaling and few-shot learning.

Statistics

Statistical analyses were performed in Python 3.9. Welch's t-test was performed, using either the Pingouin[83] (v0.5.3) or scipy.stats (v1.7.3) packages, to compare between experimental groups, so as to not assume equal variances between the populations. To test for equality of variances, F-tests were run using a custom script. Multiple t-tests (Extended Data Fig 5) were corrected for by adjusting p-values with Bonferroni's correction. All statistical tests were two-tailed. N's reported represent individual animals or, in the case of simulations, the number of iterations. Full statistical test information is presented in Extended Data Table 1. Time courses of the cumulative sum or average of the lick and/or dopamine data are presented as mean ± SEM. Bar graphs are presented as mean ± SEM with individual animal data points. Results were considered significant at an alpha of 0.05. * denotes $p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$; ns (non-significant) denotes $p > 0.05$.

References

1. Steinberg, E. E. *et al.* A causal link between prediction errors, dopamine neurons and learning. *Nat. Neurosci.* **16**, 966–973 (2013).
2. Lee, K. *et al.* Temporally restricted dopaminergic control of reward-conditioned movements. *Nat. Neurosci.* **23**, 209–216 (2020).
3. Jeong, H. *et al.* Mesolimbic dopamine release conveys causal associations. *Science* **378**, eabq6740 (2022).
4. Maes, E. J. P. *et al.* Causal evidence supporting the proposal that dopamine transients function as temporal difference prediction errors. *Nat. Neurosci.* **23**, 176–178 (2020).
5. Schultz, W., Dayan, P. & Montague, P. R. A Neural Substrate of Prediction and Reward. *Science* **275**, 1593–1599 (1997).
6. Sutton, R. S. & Barto, A. G. Time-derivative models of pavlovian reinforcement. (1990).
7. Sutton, R. S. & Barto, A. G. A temporal-difference model of classical conditioning. in *Proceedings of the ninth annual conference of the cognitive science society* 355–378 (Seattle, WA, 1987).
8. Cohen, J. Y., Haesler, S., Vong, L., Lowell, B. B. & Uchida, N. Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature* **482**, 85–88 (2012).
9. Schultz, W., Dayan, P. & Montague, P. R. A Neural Substrate of Prediction and Reward. *Science* **275**, 1593–1599 (1997).
10. Mohebi, A. *et al.* Dissociable dopamine dynamics for learning and motivation. *Nature* **570**, 65–70 (2019).
11. Hart, A. S., Rutledge, R. B., Glimcher, P. W. & Phillips, P. E. M. Phasic Dopamine Release in the Rat Nucleus Accumbens Symmetrically Encodes a Reward Prediction Error Term. *J. Neurosci.* **34**, 698–704 (2014).
12. Day, J. J., Roitman, M. F., Wightman, R. M. & Carelli, R. M. Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens. *Nat. Neurosci.* **10**, 1020–1028 (2007).
13. Phillips, P. E. M., Stuber, G. D., Heien, M. L. A. V., Wightman, R. M. & Carelli, R. M. Subsecond dopamine release promotes cocaine seeking. *Nature* **422**, 614–618 (2003).
14. Bayer, H. M. & Glimcher, P. W. Midbrain Dopamine Neurons Encode a Quantitative Reward Prediction Error Signal. *Neuron* **47**, 129–141 (2005).
15. Kim, H. R. *et al.* A Unified Framework for Dopamine Signals across Timescales. *Cell* **183**, 1600-1616.e25 (2020).
16. Hamid, A. A., Frank, M. J. & Moore, C. I. Wave-like dopamine dynamics as a mechanism for spatiotemporal credit assignment. *Cell* **184**, 2733-2749.e16 (2021).
17. Rescorla, R. A. & Wagner, A. R. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. in *Classical Conditioning II* (eds. Black, A. H. & Prokasy, W. F.) 64–99 (Appleton-Century-Crofts, 1972).
18. Shea, C. H., Lai, Q., Black, C. & Park, J.-H. Spacing practice sessions across days benefits the learning of motor skills. *Hum. Mov. Sci.* **19**, 737–760 (2000).
19. Terrace, H. S., Gibbon, J., Farrell, L. & Baldock, M. D. Temporal factors influencing the acquisition and maintenance of an autoshaped keypeck. *Anim. Learn. Behav.* **3**, 53–62 (1975).
20. Menzel, R., Manz, G., Menzel, R. & Greggers, U. Massed and Spaced Learning in Honeybees: The Role of CS, US, the Intertrial Interval, and the Test Interval. *Learn. Mem.* **8**, 198–208 (2001).
21. Beck, C. D. O., Schroeder, B. & Davis, R. L. Learning Performance of Normal and Mutant Drosophila after Repeated Conditioning Trials with Discrete Stimuli. *J. Neurosci.* **20**, 2944–2953 (2000).
22. Mauelshagen, J., Sherff, C. M. & Carew, T. J. Differential Induction of Long-Term Synaptic Facilitation by Spaced and Massed Applications of Serotonin at Sensory Neuron Synapses of Aplysia californica. *Learn. Mem.* **5**, 246–256 (1998).
23. Reynolds, B. The acquisition of a trace conditioned response as a function of the magnitude of the stimulus trace. *J. Exp. Psychol.* **35**, 15–30 (1945).

24. Fanselow, M. S. & Tighe, T. J. Contextual conditioning with massed versus distributed unconditional stimuli in the absence of explicit conditional stimuli. *J. Exp. Psychol. Anim. Behav. Process.* **14**, 187–199 (1988).
25. Ebbinghaus, H. *Memory: A contribution to experimental psychology*. viii, 128 (Teachers College Press, 1913). doi:10.1037/10011-000.
26. Smolen, P., Zhang, Y. & Byrne, J. H. The right time to learn: mechanisms and optimization of spaced learning. *Nat. Rev. Neurosci.* **17**, 77–88 (2016).
27. Lattal, K. M. Trial and intertrial durations in Pavlovian conditioning: Issues of learning and performance. *J. Exp. Psychol. Anim. Behav. Process.* **25**, 433–450 (1999).
28. Holland, P. C. Trial and intertrial durations in appetitive conditioning in rats. *Anim. Learn. Behav.* **28**, 121–135 (2000).
29. Sunsay, C. & Bouton, M. E. Analysis of a trial-spacing effect with relatively long intertrial intervals. *Learn. Behav.* **36**, 104–115 (2008).
30. Ward, R. D. *et al.* CS Informativeness Governs CS-US Associability. *J. Exp. Psychol. Anim. Behav. Process.* **38**, 217–232 (2012).
31. Sunsay, C., Stetson, L. & Bouton, M. E. Memory priming and trial spacing effects in Pavlovian learning. *Anim. Learn. Behav.* **32**, 220–229 (2004).
32. Gottlieb, D. A. Is the number of trials a primary determinant of conditioned responding? *J. Exp. Psychol. Anim. Behav. Process.* **34**, 185–201 (2008).
33. Wimmer, G. E., Li, J. K., Gorgolewski, K. J. & Poldrack, R. A. Reward Learning over Weeks Versus Minutes Increases the Neural Representation of Value in the Human Brain. *J. Neurosci. Off. J. Soc. Neurosci.* **38**, 7649–7666 (2018).
34. Gibbon, J. & Balsam, P. Spreading associations in time. in *Autoshaping and conditioning theory* (eds. Locurto, C. M., Terrace, H. S. & Gibbon, J.) 219–253 (New York: Academic, 1981).
35. Gallistel, C. R. & Gibbon, J. Time, rate, and conditioning. *Psychol. Rev.* **107**, 289 (2000).
36. Gallistel, C. R., Craig, A. R. & Shahan, T. A. 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).
37. Ludvig, E. A., Sutton, R. S. & Kehoe, E. J. Stimulus Representation and the Timing of Reward-Prediction Errors in Models of the Dopamine System. *Neural Comput.* **20**, 3034–3054 (2008).
38. Thrailkill, E. A., Todd, T. P. & Bouton, M. E. Effects of conditioned stimulus (CS) duration, intertrial interval, and I/T ratio on appetitive Pavlovian conditioning. *J. Exp. Psychol. Anim. Learn. Cogn.* **46**, 243–255 (2020).
39. Waelti, P., Dickinson, A. & Schultz, W. Dopamine responses comply with basic assumptions of formal learning theory. *Nature* **412**, 43–48 (2001).
40. Cohen, J. Y., Haesler, S., Vong, L., Lowell, B. B. & Uchida, N. Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature* **482**, 85–88 (2012).
41. Namboodiri, V. M. K. *et al.* 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).
42. Gallistel, C. R. & Papachristos, E. B. Number and time in acquisition, extinction and recovery. *J. Exp. Anal. Behav.* **113**, 15–36 (2020).
43. Gallistel, C. R., Fairhurst, S. & Balsam, P. The learning curve: Implications of a quantitative analysis. *Proc. Natl. Acad. Sci.* **101**, 13124–13131 (2004).
44. Vega-Villar, M., Horvitz, J. C. & Nicola, S. M. NMDA receptor-dependent plasticity in the nucleus accumbens connects reward-predictive cues to approach responses. *Nat. Commun.* **10**, 4429 (2019).
45. Moore, S. & Kuchibhotla, K. V. Slow or sudden: Re-interpreting the learning curve for modern systems neuroscience. *IBRO Neurosci. Rep.* **13**, 9–14 (2022).
46. Pearce, J. M. & Hall, G. A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychol. Rev.* **87**, 532–552 (1980).
47. Mackintosh, N. J. Overshadowing and stimulus intensity. *Anim. Learn. Behav.* **4**, 186–192 (1976).

48. Aitken, T. J., Greenfield, V. Y. & Wassum, K. M. Nucleus accumbens core dopamine signaling tracks the need-based motivational value of food-paired cues. *J. Neurochem.* **136**, 1026–1036 (2016).
49. Day, J. J., Roitman, M. F., Wightman, R. M. & Carelli, R. M. Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens. *Nat. Neurosci.* **10**, 1020–1028 (2007).
50. Menegas, W., Babayan, B. M., Uchida, N. & Watabe-Uchida, M. Opposite initialization to novel cues in dopamine signaling in ventral and posterior striatum in mice. *eLife* **6**, e21886 (2017).
51. Flagel, S. B. *et al.* A selective role for dopamine in stimulus–reward learning. *Nature* **469**, 53–57 (2011).
52. Coddington, L. T. & Dudman, J. T. The timing of action determines reward prediction signals in identified midbrain dopamine neurons. *Nat. Neurosci.* **21**, 1563–1573 (2018).
53. Fanselow, M. S. Neural organization of the defensive behavior system responsible for fear. *Psychon. Bull. Rev.* **1**, 429–438 (1994).
54. McKay, Leah, Hunninck, Louis & Sheriff, Michael J. A Field-Based Adaptation of the Classic Morris Water Maze to Assess Learning and Memory in a Free-Living Animal. *Anim. Behav. Cogn.* **9**, 396–407 (2022).
55. Bruce, H. M. An Exteroceptive Block to Pregnancy in the Mouse. *Nature* **184**, 105–105 (1959).
56. Klee, J. L., Souza, B. C. & Battaglia, F. P. Learning differentially shapes prefrontal and hippocampal activity during classical conditioning. *eLife* **10**, e65456 (2021).
57. Meister, M. Learning, fast and slow. *Curr. Opin. Neurobiol.* **75**, 102555 (2022).
58. Rosenberg, M., Zhang, T., Perona, P. & Meister, M. Mice in a labyrinth show rapid learning, sudden insight, and efficient exploration. *eLife* **10**, e66175 (2021).
59. Ludvig, E. A., Mirian, M. S., Kehoe, E. J. & Sutton, R. S. Associative Learning from Replayed Experience. 100800 Preprint at https://doi.org/10.1101/100800 (2017).
60. Nguyen, N. D. *et al.* Cortical reactivations predict future sensory responses. 2022.11.14.516421 Preprint at https://doi.org/10.1101/2022.11.14.516421 (2022).
61. Carr, M. F., Jadhav, S. P. & Frank, L. M. Hippocampal replay in the awake state: a potential physiological substrate of memory consolidation and retrieval. *Nat. Neurosci.* **14**, 147–153 (2011).
62. Sugden, A. U. *et al.* Cortical reactivations of recent sensory experiences predict bidirectional network changes during learning. *Nat. Neurosci.* **23**, 981–991 (2020).
63. Diba, K. & Buzsáki, G. Forward and reverse hippocampal place-cell sequences during ripples. *Nat. Neurosci.* **10**, 1241–1242 (2007).
64. Foster, D. J. & Wilson, M. A. Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature* **440**, 680–683 (2006).
65. Gupta, A. S., van der Meer, M. A. A., Touretzky, D. S. & Redish, A. D. Hippocampal Replay Is Not a Simple Function of Experience. *Neuron* **65**, 695–705 (2010).
66. Gillespie, A. K. *et al.* Hippocampal replay reflects specific past experiences rather than a plan for subsequent choice. *Neuron* **109**, 3149-3163.e6 (2021).
67. Pietrock, C. *et al.* Pupil dilation as an implicit measure of appetitive Pavlovian learning. *Psychophysiology* **56**, e13463 (2019).
68. Yamada, K. & Toda, K. Pupillary dynamics of mice performing a Pavlovian delay conditioning task reflect reward-predictive signals. *Front. Syst. Neurosci.* **16**, (2022).
69. Finke, J. B., Roesmann, K., Stalder, T. & Klucken, T. Pupil dilation as an index of Pavlovian conditioning. A systematic review and meta-analysis. *Neurosci. Biobehav. Rev.* **130**, 351–368 (2021).
70. Lutas, A., Fernando, K., Zhang, S. X., Sambangi, A. & Andermann, M. L. History-dependent dopamine release increases cAMP levels in most basal amygdala glutamatergic neurons to control learning. *Cell Rep.* **38**, 110297 (2022).
71. Mohebi, A. *et al.* Dissociable dopamine dynamics for learning and motivation. *Nature* **570**, 65–70 (2019).
72. Kutlu, M. G. *et al.* Dopamine release in the nucleus accumbens core signals perceived saliency. *Curr. Biol.* **31**, 4748-4761.e8 (2021).

73. Hamid, A. A. *et al.* Mesolimbic dopamine signals the value of work. *Nat. Neurosci.* **19**, 117–126 (2016).
74. Sharpe, M. J. *et al.* Dopamine transients do not act as model-free prediction errors during associative learning. *Nat. Commun.* **11**, 106 (2020).
75. Kalmbach, A. *et al.* Dopamine encodes real-time reward availability and transitions between reward availability states on different timescales. *Nat. Commun.* **13**, 3805 (2022).
76. Carter, F. *et al.* Does phasic dopamine release cause policy updates? 2022.08.08.502043 Preprint at https://doi.org/10.1101/2022.08.08.502043 (2022).
77. Goh, W. Z., Ursekar, V. & Howard, M. W. Predicting the future with a scale-invariant temporal memory for the past. *ArXiv210110953 Cs Q-Bio* (2021).
78. Shankar, K. H. & Howard, M. W. A scale-invariant internal representation of time. *Neural Comput.* **24**, 134–193 (2012).
79. Namboodiri, V. M. K. How do real animals account for the passage of time during associative learning? *Behav. Neurosci.* **136**, 383–391 (2022).
80. Lerner, T. N. *et al.* Intact-Brain Analyses Reveal Distinct Information Carried by SNc Dopamine Subcircuits. *Cell* **162**, 635–647 (2015).
81. Clark, J. J., Collins, A. L., Sanford, C. A. & Phillips, P. E. M. Dopamine Encoding of Pavlovian Incentive Stimuli Diminishes with Extended Training. *J. Neurosci.* **33**, 3526–3532 (2013).
82. Trujillo-Pisanty, I., Conover, K., Solis, P., Palacios, D. & Shizgal, P. 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).
83. Vallat, R. Pingouin: statistics in Python. *J. Open Source Softw.* **3**, 1026 (2018).

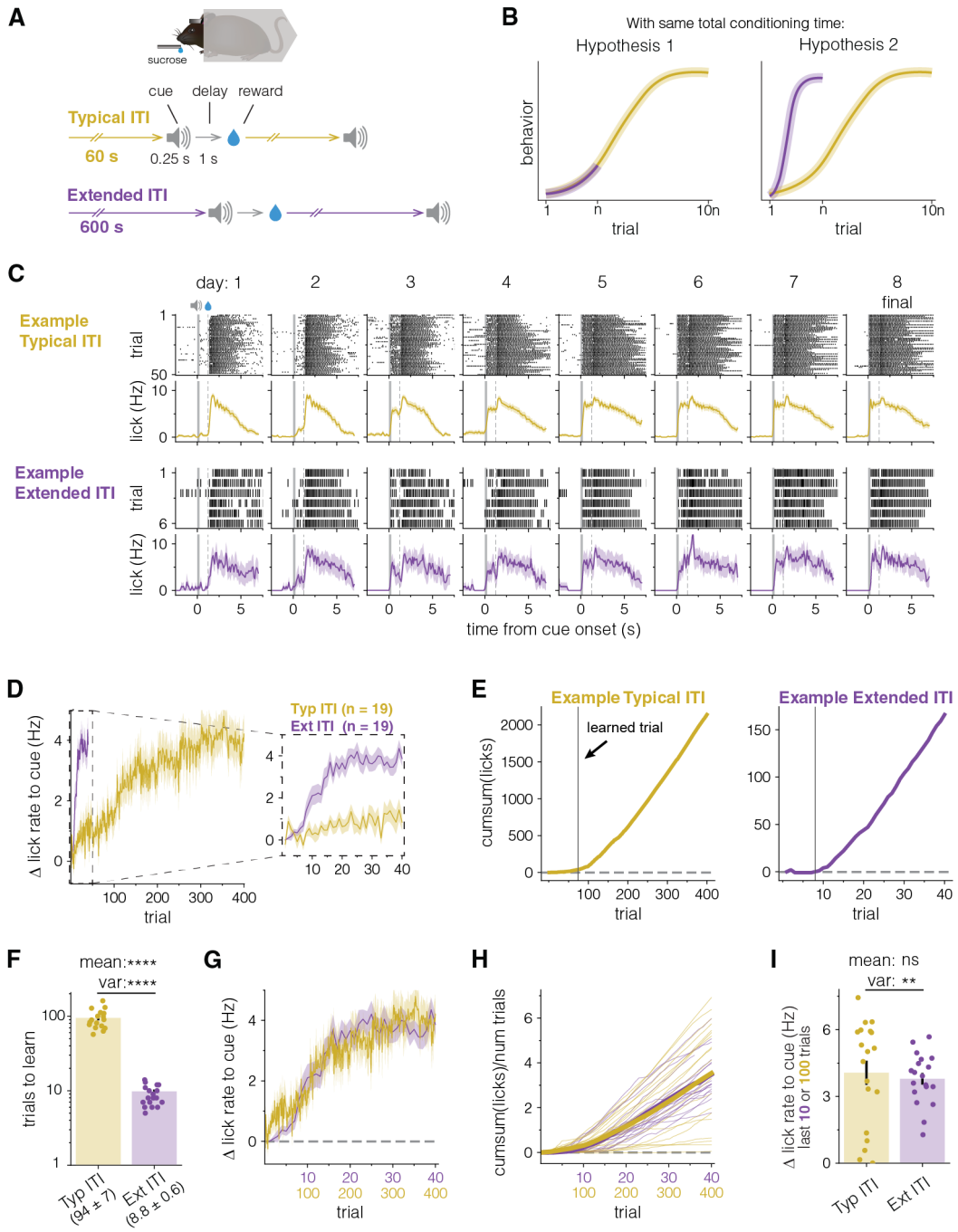


Figure 1. Temporal scaling in behavioral learning.

A. Schematic of experimental setup. Head-fixed mice were divided into two groups that were each presented with identical cue-reward pairing trials, but each group differed in the length of time between cue-reward presentations (i.e., the inter-trial interval or ITI). Trials consisted of an auditory cue (0.25 s, 12 kHz) followed by a 1 s delay before a drop of sucrose solution (15% w/v, ~2.5 μL) was delivered through a spout in front of the animal. The Typical ITI group had an average ITI of 60 s, while the Extended ITI group had an average ITI of 600 s. The total conditioning time per day was kept roughly constant resulting in 50 trials a day for Typical ITI mice, and 6 trials a day for Extended ITI mice (this accounts for a fixed 3 s reward consumption period; see Methods). Mice were run for 8 days.

B. Illustration of two hypothetical experimental outcomes. Learning curves display the possible relationship between Typical ITI and Extended ITI group learning rates as a function of trial number. Hypothesis 1 is based on standard "trial-based" neuroscience models used to explain reward learning, including TDRL. These posit that as long as the ITI is sufficiently longer than the trial duration (as is the case in both experimental groups), there will be no difference in learning between groups presented with different ITIs (See Extended Data Fig 1). Because total conditioning time is kept roughly constant between groups, the Extended ITI mice will show less evidence of learning than Typical ITI mice due to ten times fewer trial experiences by the end of conditioning. Hypothesis 2 is that previous demonstrations of faster learning in "spaced learning" applies even when the ITI is much longer than trial duration (ratio of 48 in Typical ITI group). Here, we present a strict version of this hypothesis in which the group that experiences 10 times fewer trials learns 10 times more per trial.

C. Example lick raster plots (upper) and lick peri-stimulus time histograms (PSTH)(lower) for one example mouse from either Typical ITI group (top, gold) or Extended ITI group (bottom, purple) showing every cue-reward delivery epoch across the full eight days of conditioning. Each column represents a single day of conditioning. Graphs are aligned to cue onset (cue on denoted by gray shading). Reward delivery is denoted by the vertical gray dashed line. Both example animals begin to show evidence of learning (an increase in licking following cue onset before reward is delivered) on day 2.
D. Extended ITI mice learn and reach asymptotic behavior in fewer trials than Typical ITI mice. **Left,** Time course showing the average change in cue-evoked lick rate (the baseline subtracted lick rate between cue onset and reward delivery, see Methods) over 40 (Ext ITI, purple, n = 19 mice) or 400 (Typ ITI, gold, n = 19 mice) cue-reward presentations. **Inset right**, Zoom in of first 40 trials for both groups. Lines represent mean across animals and shaded area represents the SEM.
E. Cumulative sum (cumsum) of cue-evoked licks across trials from the same example mice as in **C**. Using the cumsum curve from each animal to determine the trial at which mice first show evidence of learning (see Methods), we found that the example Typical ITI mouse (left, gold) learns at trial 74, while Extended ITI group (right, purple) learns at trial 8 (i.e., "few shot" reward learning). Learned trial is denoted by the solid black vertical line.
F. Extended ITI mice learn in about ten times fewer trials than Typical ITI mice. Bar height represents mean trial at which mice show evidence of learning for Typ ITI group (left, gold, n = 17) and Ext ITI group (right, purple, n = 19), plotted on a log scale. Error bar represents SEM. Circles represent individual mice. Values under labels represent mean ±SEM. Two mice that did not show evidence of learning are excluded from comparison (Extended Data Fig 3; see Methods). **** $p < 0.0001$, Welch's t-test, F-test.
G & H. On average, learning between groups progresses similarly as a function of total conditioning time, and thus it scales with the ratio of ITIs. Learning rate of one Ext ITI trial is similar to that of ten Typ ITI trials. **G.** Mean cue-evoked lick rates for Ext ITI and Typ ITI groups across scaled trial numbers (same data as in D), showing that the Ext ITI group learns ten times higher per experience compared to the Typ ITI group. **H.** Cumsum of cue-evoked licks plotted on the same scaled x-axis. Thick lines represent group means and individual lines represent individual animals. There is much higher individual variability in the Typ ITI group compared to the Ext ITI group (quantified in **I**).
I. Asymptotic cue-evoked lick rates have similar group means, but different variances. Bars represent mean cue-evoked lick rates during trials 301–400 (Typ ITI) or trials 31-40 (Ext ITI). Error bars represent SEMs and circles represent individual mice. ns: not significant, Welch's t-test; **$p<0.01$, F-test.
See Extended Data Table 1 for full statistical test details. All error bars and error shading throughout manuscript represent SEM.

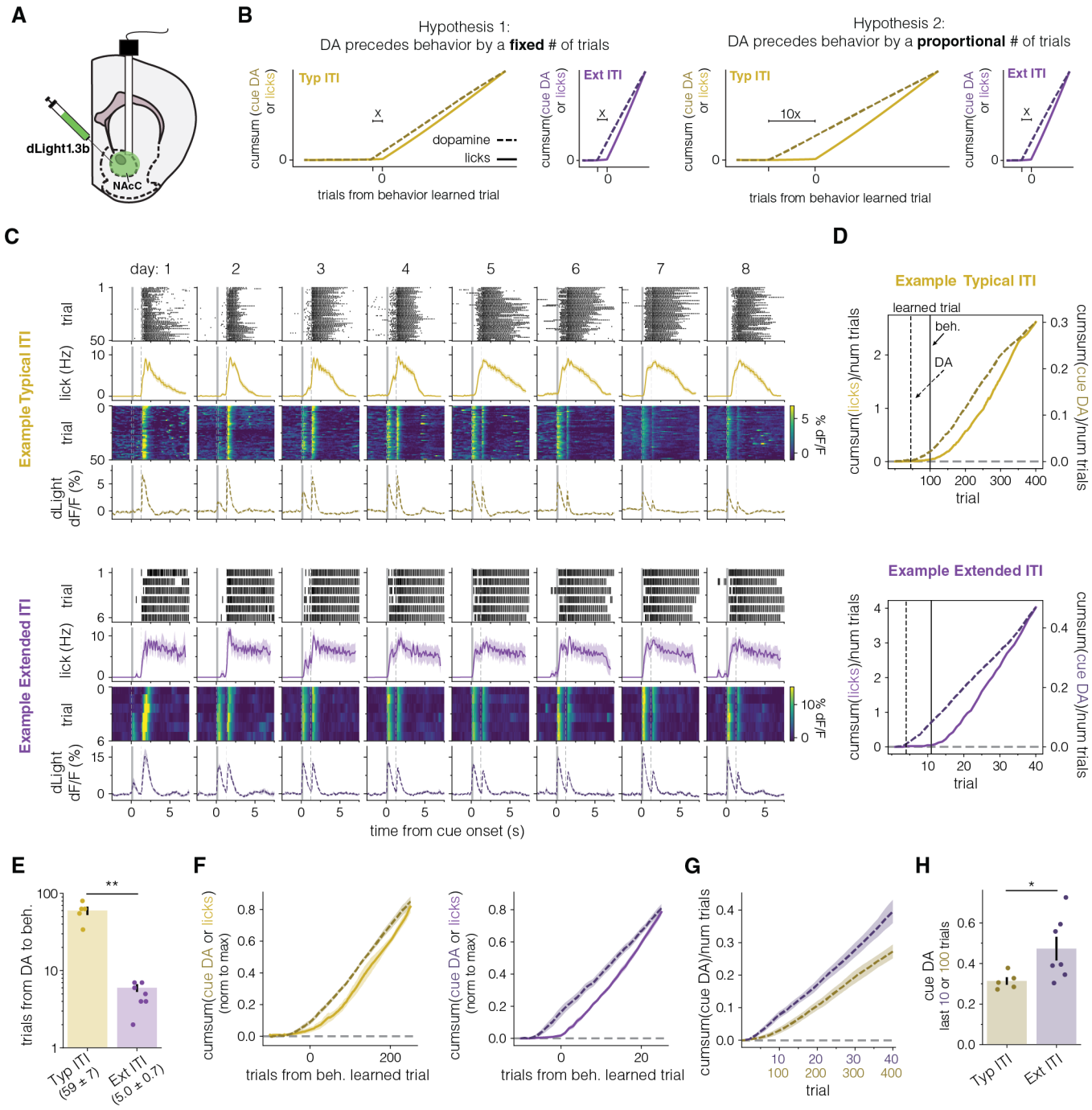


Figure 2. Temporal scaling in mesolimbic dopaminergic learning.

A. Schematic of mesolimbic dopamine measurements (Methods).

B. Diagrams of 2 hypothetical relationships between cue-evoked dopaminergic learning and behavioral learning in Typ ITI and Ext ITI mice. Hypothesis 1 is that the development of cue-evoked dopamine precedes behavioral learning by a fixed number of trials in both Typ ITI and Ext ITI groups, suggesting a one-to-one evolution of dopaminergic and behavioral learning. This hypothesis is based on models of dopamine function that posit a tight relationship between cue-evoked dopamine and behavioral learning whereby the development of dopaminergic cue responses directly drives learning. Hypothesis 2 is that the development of cue-evoked dopamine precedes behavioral learning by a number of trials proportional to the number needed for behavioral learning. Because Ext ITI mice learn in ten times fewer trials than Typ ITI mice (Fig 1F), this hypothesis predicts that cue-evoked dopamine will precede behavioral learning by ten times fewer trials in Ext ITI mice.

C. Example lick raster plots (upper row), lick PSTH (2nd row), heatmap of dopamine responses on each trial (3rd row) and average dopamine response for the day (lower row) for one example mouse from either Typical ITI group (top, gold) or Extended ITI group (bottom, purple) during cue and reward presentation across 8 days of conditioning. Lick data presented as in Fig 1C. Dopamine signals plotted as % dF/F. Graphs aligned to cue onset (cue on denoted by gray shading). Reward delivery is denoted by vertical gray dashed line.

D. Cumsum of cue-evoked licks (solid, lighter, left axis) or of cue-evoked dopamine (dashed, darker, right axis) for the same example mice as in C. Both lick and cue dopamine values were divided by total trial number to display average responses across conditioning. Before taking the cumsum, cue-evoked dopamine responses were normalized by max reward responses (see Methods). Cumsum curves were used to determine the trial at which cue-evoked dopamine and cue-evoked licking emerge ("learned trial", see Methods). Solid vertical lines represent learned behavior trial and dashed vertical lines represent dopamine learned trial.
E. On average, DA cue responses develop 59 trials before the emergence of cue-evoked licking in Typ ITI mice and 5 trials before in Ext ITI mice. Bars represent mean number of trials between dopamine and behavior learned trials. Error bars show SEM. Circles represent individual mice. Values under labels represent mean ±SEM. **$p<0.01$. Welch's t-test.
F. Mean cumsum of cue-evoked licking (solid) and dopamine (dashed) for Typ ITI (left, gold, n = 5 animals) and Ext ITI (right, purple, n = 7 animals) mice. Data were normalized by each animal's final trial of conditioning and aligned to their learned trial before averaging. Lines represent means, and shading represents SEM.
G. Mean cumsum of reward normalized cue-evoked dopamine responses in Typ ITI (gold) and Ext ITI (purple) mice. Cumsum curves normalized by number of trials to account for differences between groups.
H. Mean asymptotic cue-evoked dopamine after learning normalized by individual max reward responses. Bars represent means for trials 301-400 (Typ ITI) or trials 31-40 (Ext ITI). Error bars represent SEM and circles represent individual mice. *$p<0.05$, Welch's t-test.

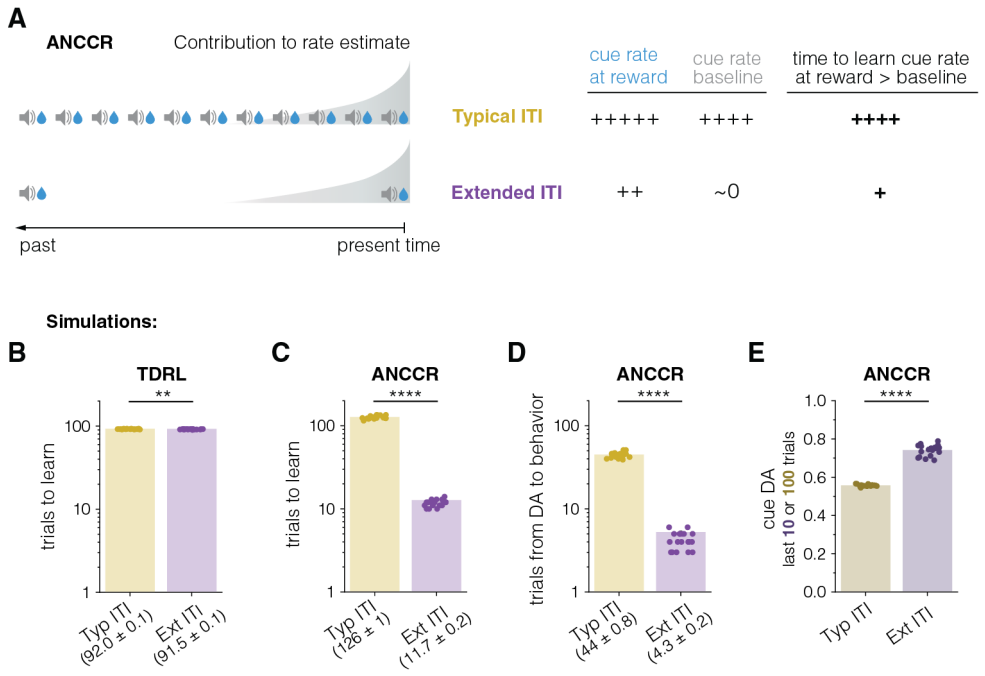
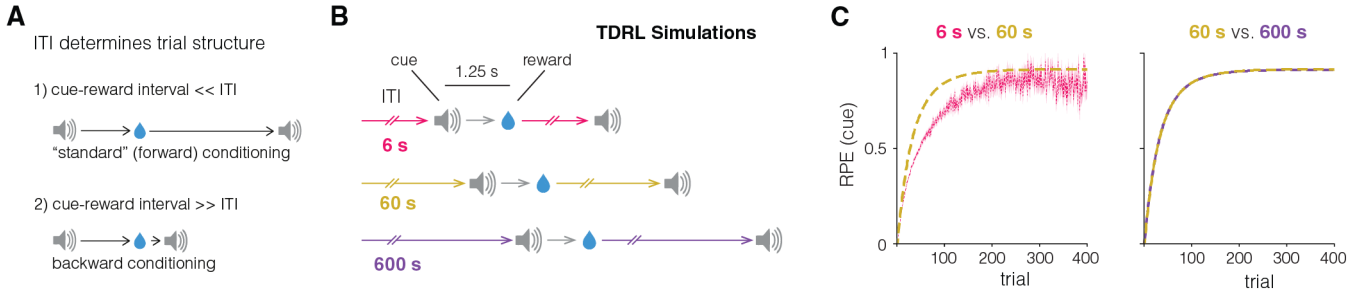


Figure 3. Simulations based on a model of retrospective causal learning capture the experimentally observed quantitative scaling relationship between inter-trial interval, dopamine, and learning.

A. Schematic showing the intuition behind the model that explains differences in per trial learning between Typ ITI and Ext ITI groups. In this model, animals learn cue-reward associations through calculation of an adjusted net contingency for causal relations (ANCCR) based on estimates of the rate of cues at time of rewards vs. the baseline rate of cues. Cue presentations evoke an exponentially decaying eligibility trace whose mean is calculated either at reward time or baseline to determine whether reward is contingent upon cue presentation. If the eligibility traces evoked by cues are of such a duration that prior cue eligibility traces have not fully decayed before presentation of subsequent cues and rewards for Typ ITI, but not Ext ITI, then learning will be slower for Typ ITI animals. Cue rate at time of reward and baseline will both be high in Typ ITI, slowing down contingency estimates relative to Ext ITI group where estimates of the cue rate at baseline will be near zero.

B. Quantification of learned trial from TDRL simulation data plotted in Extended Data Fig 1C for 60 s (gold, n = 20 iterations) and 600 s (purple, n = 20 iterations) ITI conditions. Bars represent mean number of trials before learning occurs plotted on a log scale. Error bars show SEM. Circles represent individual iterations. Values under labels represent mean ±SEM. **p<0.01, Welch's t-test.

C – E. ANCCR simulations of cue-reward learning using same trial parameters and ITIs as mouse experiments capture the experimental observations that conditioning with an Extended ITI (600 s, purple, n = 20 iterations) leads to learning in ten times fewer trials (**C**), ten times fewer trials between the development of cue-evoked dopamine and learning (**D**), and greater asymptotic cue-evoked dopamine responses (**E**) compared to conditioning with a Typical ITI (60 s, gold, n = 20 iterations). TDRL models of dopamine do not capture any of these effects. Bars represent means. Error bars show SEM. Circles represent individual iterations. Values under labels represent mean ±SEM. ****p<0.0001, Welch's t-tests.

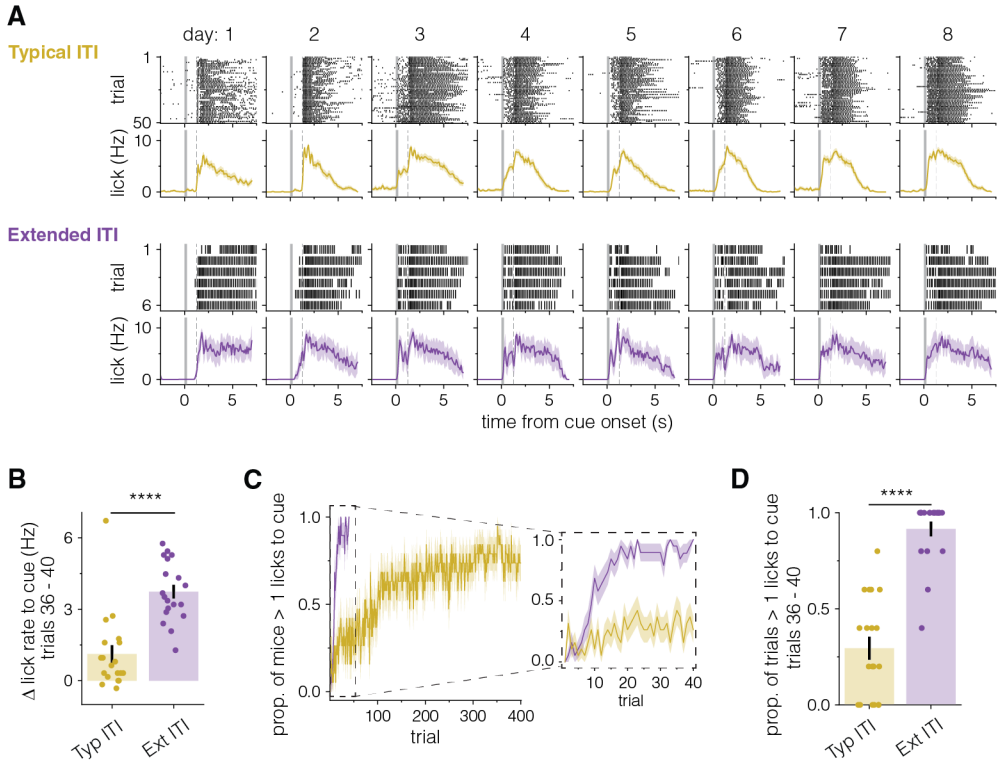


Extended Data Figure 1. Temporal difference reinforcement learning (TDRL) algorithms predict no difference in learning as a function of inter-trial interval as long as the inter-trial interval is much longer than trial duration.

A. Illustration of how inter-trial interval (ITI) can impact learning. The ITI is an important parameter in cue-reward conditioning because its relationship to the cue-reward interval (trial duration) determines the overall structure of cue-reward experiences. In standard conditioning paradigms where the ITI is much longer than the trial duration (1, top), associations that the cue predicts the reward are commonly learned. In an extreme counter-example, however, if the ITI is much shorter than the same trial duration, then the association learned is that the reward predicts the cue, as in backward conditioning, which is known to produce either little or even negative conditioning. So, the ITI must impact learning. This is a simple intuitive explanation for prior observations that "spaced learning" is more effective than "massed learning". This explanation assumes that once the ITI is much longer than the trial duration, there is no effect of the ITI on conditioning, consistent with TDRL models (**B, C**).

B. Diagram of cue-reward learning simulations for three conditions in which trial duration is identical and ITI is varied across conditions. Trials consists of a 1.25 s long cue to reward delay (trial duration). In one condition, ITI is kept short at 6 s (ITI/trial ratio of 4.8). In another condition, ITI is kept to a typical value (60 s) but with a high ITI/trial ratio (48). In a third condition, ITI is extended in duration at 600 s (ITI/trial ratio of 480).

C. Temporal difference reinforcement learning (TDRL) simulations using the microstimulus model that naturally accounts for the ITI predict that conditioning with a 60 s ITI (ITI/trial = 48) will lead to faster learning than conditioning with a 6 s ITI (ITI/trial = 4.8), but that conditioning with a 600 s ITI (ITI/trial = 480) will lead to similar learning and behavior as conditioning with a 60 s ITI.

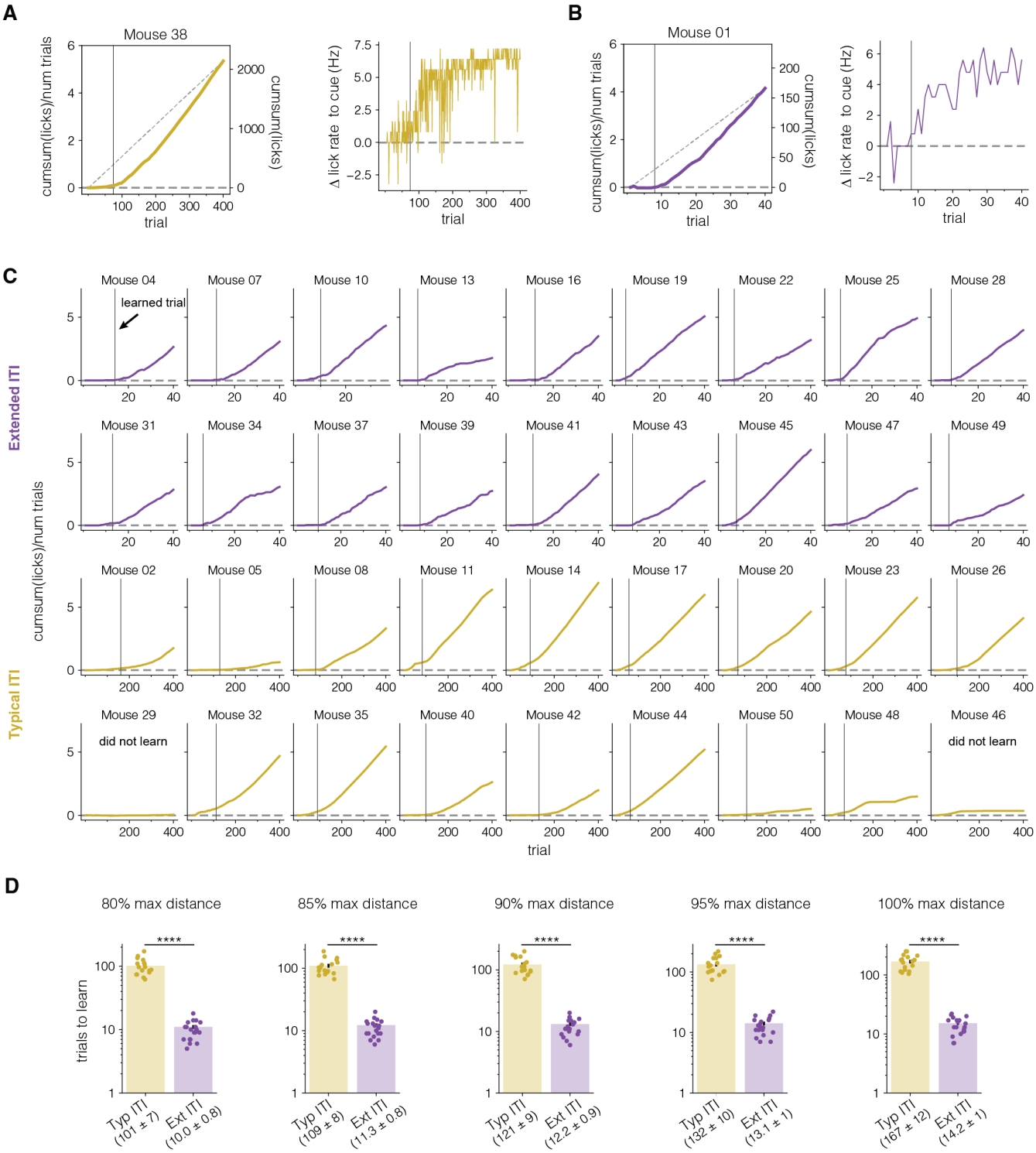


Extended Data Figure 2. Ext ITI mice show evidence of learning in significantly fewer trials than Typ ITI mice.

A. More lick raster and PSTH plots from individual Typ ITI and Ext ITI example mice as in Fig 1C.

B. Average change in cue-evoked lick rate for trials 36 - 40 (time course in Fig 1D). Ext ITI show significantly more licking to cue in this period than Typ ITI mice ****p< 0.0001, Welch's t-test.

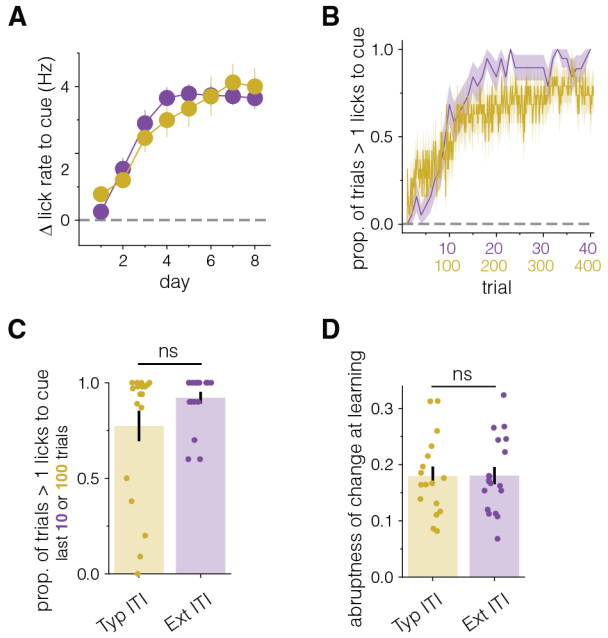
C & D. Ext ITI mice show asymptotic responding to the cue in fewer trials than Typ ITI mice. **C**. **Left,** Time course showing the proportion of mice on each trial with more than one cue-evoked lick over 40 (Ext ITI, purple, n = 19) or 400 (Typ ITI, gold, n = 19) trials. **Inset, right**, Zoom in of first 40 trials for both groups. Lines represent mean across all animals and shaded area represents the SEM. **D**. Bar height represents proportion of trials in which animals responded to cue with more than one lick between trials 36 and 40. Error bars represent SEMs, and circles represent individual animals ****p< 0.0001, Welch's t-test.



Extended Data Figure 3. The cumulative sum of cue-evoked lick behavior allows for determination of learned trial in individual animals.

A & B. Cumsum and trial-by-trial plots of cue-evoked licking behavior of same example mice as in Fig 1C & 1E. **Left,** Cumsum plot showing diagonal line used to calculate learned trials. Right y-axis shows the total cumsum of all cue-evoked licks as in Fig 1E. Left y-axis shows cumsum of cue-evoked licks divided by total trial number to allow comparisons across groups that experienced a different number of cue-reward pairings. Trial normalized cumsum values represent the mean number of cue-evoked licks over all previous trials. Solid vertical line represents the calculated learned trial, the first trial at which animals show evidence of learning. **Right,** The cue-evoked change in lick rate plotted for the same individual example mouse. Note how the vertical line representing the learned trial, which corresponds to the point on cumsum plots where the cumsum curve takes off from the x-axis, captures the trial at which cue-evoked changed in lick rates become consistently positive.

C. Cumsum plots for all remaining mice included in behavior analysis plotted with trial normalized units (see **A** & **B**). Vertical line represents calculated learned trial. Animals which did not meet learning criteria (see Methods) are noted, and no vertical line is drawn. These animals were excluded from comparison of learned trials between groups.
D. For analysis, learned trial was calculated as the first trial that fell within 75% of the maximum distance from a diagonal drawn from the point on the cumsum curve at trial 1 through trial 40 or 400. 75% of maximum distance was chosen rather than the overall maximum distance (which would be the largest inflection point in the curve) to account for variability in post-learning behavior that occasionally caused the maximum distance from the diagonal to be at a point after a mouse has consistently licked to the cue for many trials. This choice did not affect our main conclusions as using 80%, 85%, 90%, 95%, or the maximum distance from the diagonal in our algorithm yielded a roughly similar result of ten times more trials needed to learn in Typ ITI mice than Ext ITI mice. Bar heights represent mean number of trials to learn, error bars represent SEMs, and circles represent individual animals plotted on a log scale. Values under labels represent mean ±SEM. ****$p<0.0001$, Welch's t-test.

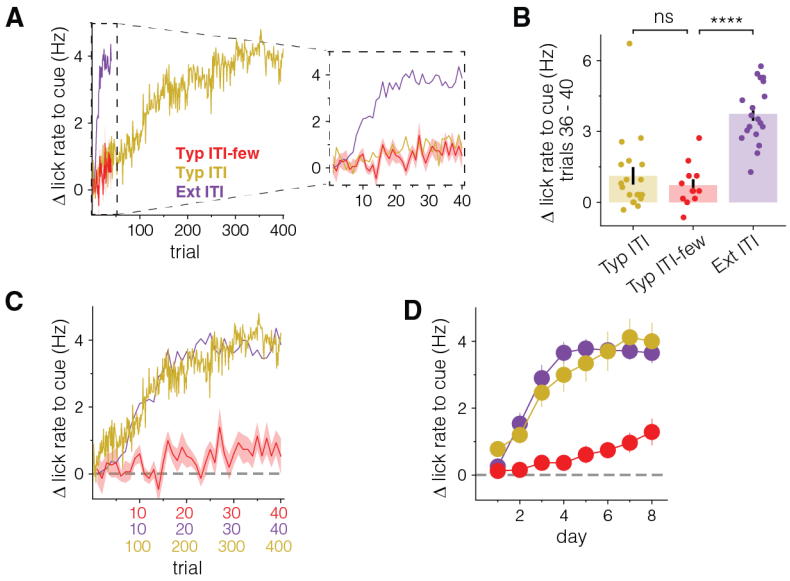


Extended Data Figure 4. Learning scales with total conditioning time or the ratio between inter-trial intervals.

A. Learning is similar between groups when trials are averaged across days. Time course of mean change in cue-evoked lick rates as a function of days of conditioning. Circles represent mean change in lick rate per day, and error bars represent SEMs.

B & C. Responding to cue scales with total conditioning time between groups and is not different at the end of conditioning. **B.** Time course showing the proportion of mice with more than one cue-evoked lick on each trial (same data as Extended Data Fig 2C) plotted on scaled x-axis units. Lines represent means per group and shading represents SEM. **C.** Bar height represents proportion of trials in which animals responded to cue with more than one lick between trials 301 – 400 (Typ ITI) or 31 – 40 (Ext ITI). Error bars represent SEMs, and circles represent individual animals. ns: not significant, Welch's t-test.

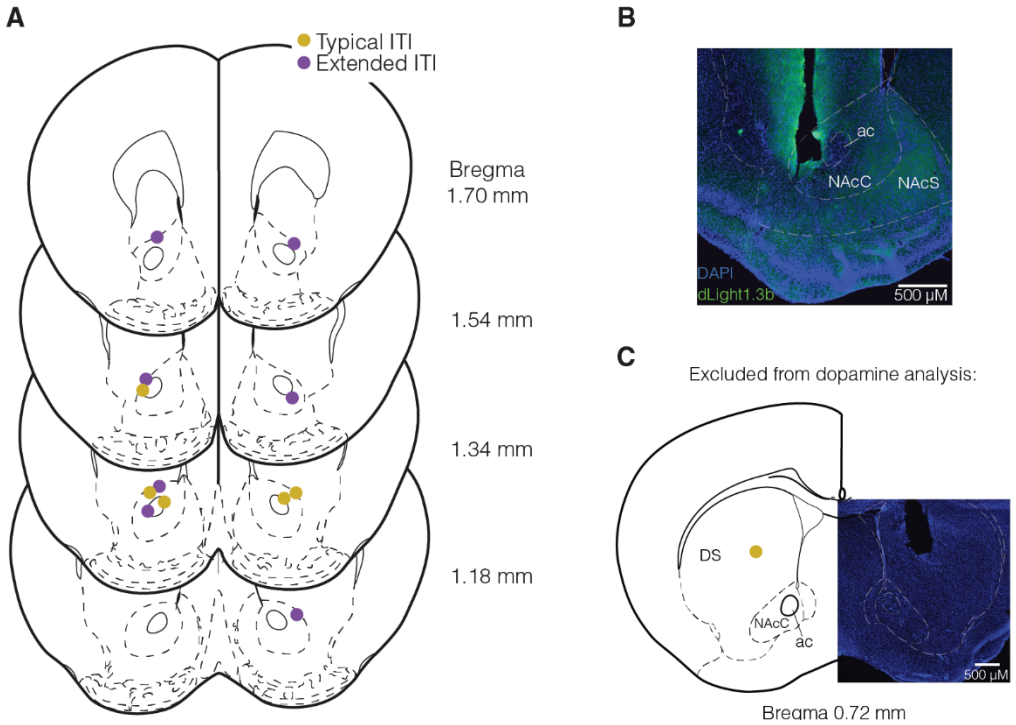
D. The abruptness of change, a measure of how quickly an animal's behavior changes at learning determined by the steepness of the lick behavior cumsum curve (see Methods), is not different between groups. Bar height represents mean abruptness of change parameter for each group. Error bars represent SEMs, and circles represent individual animals. ns: not significant, Welch's t-test.



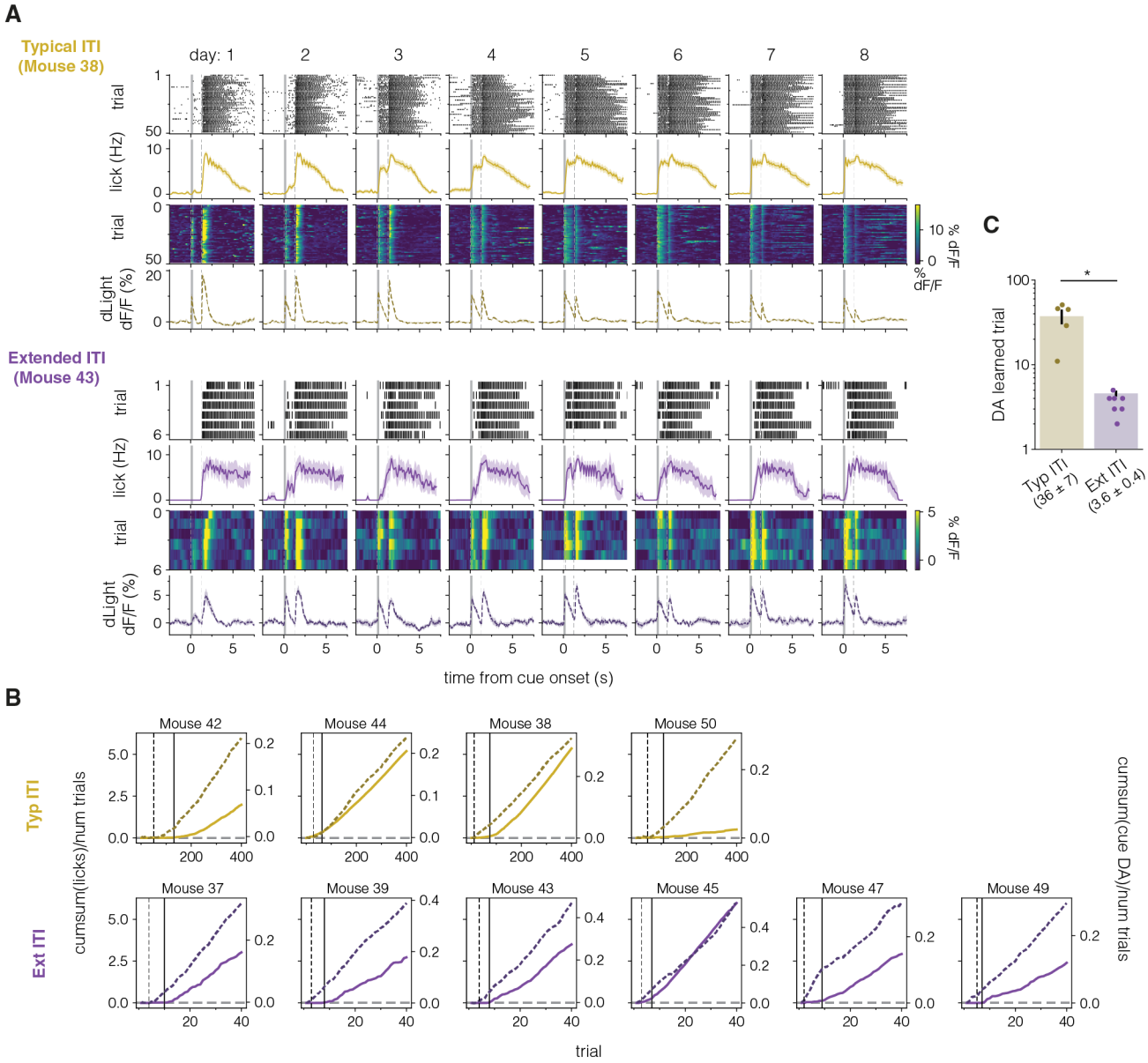
Extended Data Figure 5. Difference in learning between Typical and Extended ITI groups is not explained by difference in number of trial experiences per day.

A & B. Mice conditioned with a Typical ITI (60 s), but only six trials a day (Typ ITI-few) learn significantly less per trial than Ext ITI mice, similar to Typ ITI mice. **A.** Time course of average change in lick rate in response to the cue for Typ ITI-few (n = 12), Typ ITI (n = 19, same data as Fig 1D), and Ext ITI (n = 19, same data as Fig 1D) mice. Typ ITI and Ext ITI time courses are shown without error for visualization purposes. **B.** Average cue-evoked licking between trials 36 and 40 across all three groups. Typ ITI-few mice show significantly less evidence of learning than Ext ITI mice and behave like Typ ITI mice. ****p<0.0001, ns: not significant; Welch's t-tests.

C&D. Typ ITI-few mice show much less cue-evoked licking at the end of conditioning as shown with ITI scaled trial units (**C**) or by average licking behavior across days (**D**). Typ ITI and Ext ITI curves are the same data as in Fig 1G and Extended Data Fig 4B.



Extended Data Figure 6. Fiber placements for dopamine measurement mice.
A. Locations of center of optical fiber tip for fiber photometry recordings from Typ ITI (gold) and Ext ITI (purple) mice.
B. Example histology from a single mouse. Blue is DAPI staining and green is dLight1.3b.
C. Fiber location from Typ ITI mouse excluded from dopamine analysis.

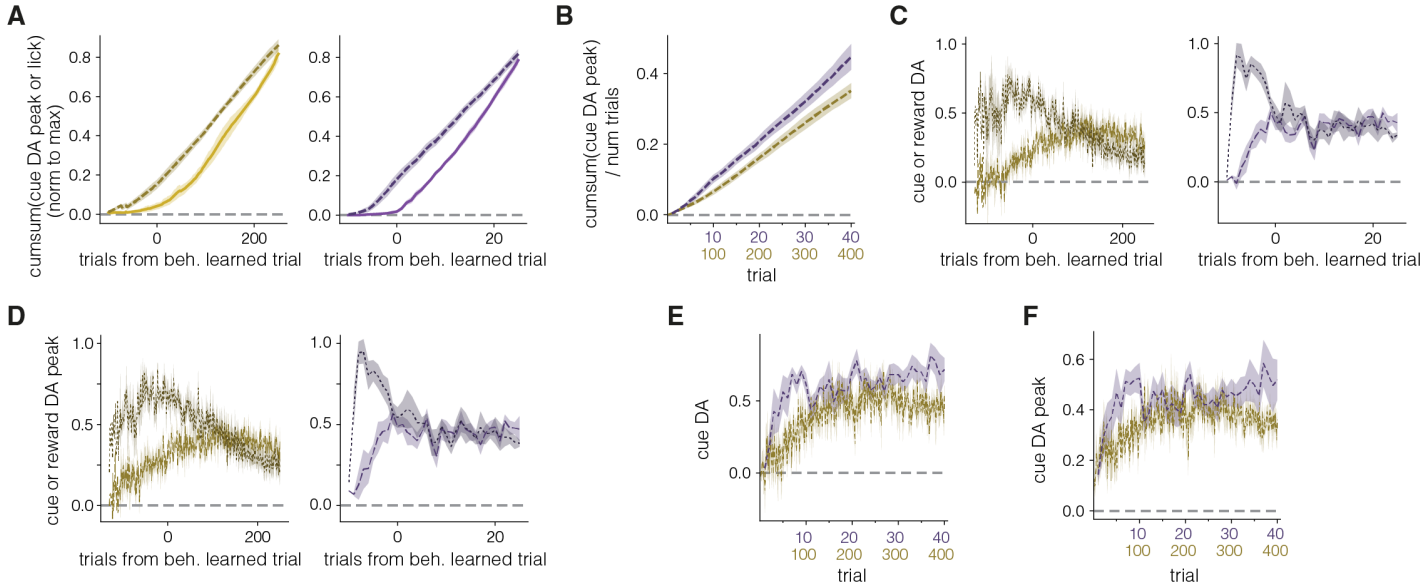


Extended Data Fig 7. Taking the cumsum of cue-evoked dopamine and licking allows for determination of trials at which dopaminergic learning and behavioral learning occur in individual mice.

A. More lick raster, lick PSTH, dopamine response by trial, and average session dopamine response plots from individual Typ ITI and Ext ITI example mice as in Fig 2C.

B. Cumsum of cue-evoked licks (solid, left axis) or of cue-evoked dopamine (dashed, right axis) for all dopamine recording mice not shown in Fig 2D. Both lick and cue evoked dopamine values were divided by total trial number to display average responses across conditioning. Before taking the cumsum, cue-evoked dopamine responses were normalized by max reward responses. Solid vertical lines represent learned behavior trial and dashed vertical lines represent dopamine learned trial.

C. Dopamine responses to cue develop in ten times fewer trials in Ext ITI mice compared to Typ ITI mice. Bar height represents mean dopamine learned trial, error bars represent SEMs, and circles represent individual mice. Values under labels represent mean ±SEM. * $p<0.05$, Welch's t-test



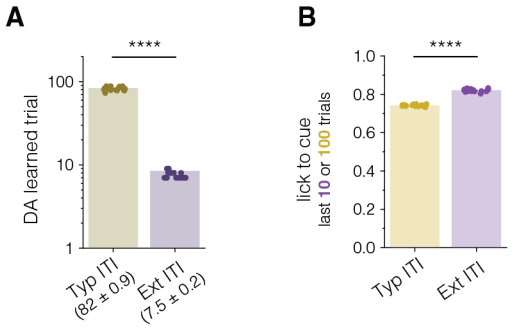
Extended Data Figure 8. Reward evoked and cue-evoked dopamine increase prior to behavioral learning.

A. Mean cumsum of cue-evoked licking (solid) and dopamine response (dashed) for Typ ITI (left, gold) and Ext ITI (right, purple) mice. Data were normalized by each animal's final trial of conditioning and aligned to their learned trial before averaging, as in Fig 2F, but using measurements of the peak of the dopamine response rather than the AUC. Lines represent means, and shading represents SEM.

B. Mean cumsum of reward normalized cue-evoked dopamine responses in Typ ITI (gold) and Ext ITI (purple) mice as in Fig 2G, but using measurements of the peak of the dopamine response instead of the AUC. Cumsum curves normalized by number of trials to account for different trial numbers between groups. Lines represent means, and shading represents SEM.

C & D. Average reward normalized AUC (**C**) or peak (**D**) cue (lighter, longer dashes) and reward (darker, smaller dashes) evoked dopamine responses aligned to each animal's behavior learned trial. Data were normalized to the average of the three maximum reward responses in each animal. Note the increase in reward response prior to behavioral learning. Lines represent means, and shading represents SEM.

E & F. Average reward normalized AUC (**E**) or peak (**F**) cue-evoked dopamine across scaled trial units. Lines represent means, and shading represents SEM.



Extended Data Figure 9. ANCCR simulations of cue-evoked licking and dopamine learned trial.
A. ANCCR simulation of the dopamine learned trial during conditioning with a Typical ITI (60 s, gold, n = 20 iterations) or Extended ITI (600 s, purple, n = 20 iterations). Bar heights represent mean response, and error bars represent SEMs. Circles represent individual iteration run. Values under labels represent mean ±SEM. ****p<0.0001, Welch's t-test.
B. ANCCR simulation of normalized lick response to cue at the end of conditioning with a Typical ITI (60 s, gold, n = 20 iterations) or Extended ITI (600 s, purple, n = 20 iterations). Bar heights represent mean response, and error bars represent SEMs. Circles represent individual iteration run. ****p<0.0001, Welch's t-test.

Extended Data Table 1 – Statistical Tests

Figure	Panel	Group (N)	Description	Test	Result	p-value (two - tailed)	Significant
Figure 1	F	**Typ ITI** (n = 17) **Ext ITI** (n = 19)	number of trials to learn	Welch's t-test	t (16.26) = -12.64	$p = 7.90 \times 10^{-10}$	****
Figure 1	F	**Typ ITI** (n = 17) **Ext ITI** (n = 19)	number of trials to learn: variance	F-test	F (16,18) = 111.2	$p = 2.18 \times 10^{-14}$	****
Figure 1	I	**Typ ITI** (n = 19) **Ext ITI** (n = 19)	change in lick rate to cue, last 10 or 100 trials	Welch's t-test	t (25.94) = 0.44	p = 0.661	ns
Figure 1	I	**Typ ITI** (n = 19) **Ext ITI** (n = 19)	change in lick rate to cue, last 10 or 100 trials: variance	F-test	F (18,18) = 4.3	p = 0.00335	**
Figure 2	E	**Typ ITI** (n = 5) **Ext ITI** (n = 7)	trials from DA to beh.	Welch's t-test	t (4.07) = 7.2	p = 0.00184	**
Figure 2	H	**Typ ITI** (n = 5) **Ext ITI** (n = 7)	cue DA, last 10 or 100 trials	Welch's t-test	t (7.16) = -2.6	p = 0.0350	*
Figure 3	B	**Typ ITI** (n = 20) **Ext ITI** (n = 20)	TDRL simulation - trials to learn	Welch's t-test	t (38) = 2.87	p = 0.00666	**
Figure 3	C	**Typ ITI** (n = 20) **Ext ITI** (n = 20)	ANCCR simulation - trials to learn	Welch's t-test	t (20.5) = 92.63	$p = 2.27 \times 10^{-28}$	****
Figure 3	D	**Typ ITI** (n = 20) **Ext ITI** (n = 20)	ANCCR simulation - trials from DA to beh.	Welch's t-test	t (22.17) = 51.3	$p = 1.53 \times 10^{-24}$	****
Figure 3	E	**Typ ITI** (n = 20) **Ext ITI** (n = 20)	ANCCR simulation - cue DA, last 10 or 100 trials: mean	Welch's t-test	t (20.11) = -28.64	$p = 8.85 \times 10^{-18}$	****
Ext. Data Figure 2	B	**Typ ITI** (n = 19) **Ext ITI** (n = 19)	change in lick rate to cue, trials 36-40	Welch's t-test	t (33.9) = -5.59	$p = 2.99 \times 10^{-6}$	****
Ext. Data Figure 2	D	**Typ ITI** (n = 19) **Ext ITI** (n = 19)	prop of trials with >1 licks to cue, trials 36 - 40	Welch's t-test	t (30.66) = -8.72	$p = 8.34 \times 10^{-10}$	****
Ext. Data Figure 3	D	**Typ ITI** (n = 17) **Ext ITI** (n = 19)	number of trials to learn (80% of max dist from diagonal)	Welch's t-test	t (16.36) = 12.46	$p = 9.06 \times 10^{-10}$	****
Ext. Data Figure 3	D	**Typ ITI** (n = 17) **Ext ITI** (n = 19)	number of trials to learn (85% of max dist from diagonal)	Welch's t-test	t (16.36) = 12.49	$p = 8.79 \times 10^{-10}$	****
Ext. Data Figure 3	D	**Typ ITI** (n = 17) **Ext ITI** (n = 19)	number of trials to learn (90% of max dist from diagonal)	Welch's t-test	t (16.27) = 11.57	$p = 2.88 \times 10^{-9}$	****
Ext. Data Figure 3	D	**Typ ITI** (n = 17) **Ext ITI** (n = 19)	number of trials to learn (95% of max dist from diagonal)	Welch's t-test	t (16.3) = 11.69	$p = 2.45 \times 10^{-9}$	****
Ext. Data Figure 3	D	**Typ ITI** (n = 17) **Ext ITI** (n = 19)	number of trials to learn (100% of max dist from diagonal)	Welch's t-test	t (16.26) = 13.13	$p = 4.51 \times 10^{-10}$	****
Ext. Data Figure 4	C	**Typ ITI** (n = 19) **Ext ITI** (n = 19)	prop. of trials with >1 licks to cue, last 10 or 100 trials	Welch's t-test	t (23.38) = -1.72	p = 0.0980	ns
Ext. Data Figure 4	D	**Typ ITI** (n = 17) **Ext ITI** (n = 19)	abruptness of change at learning	Welch's t-test	t (33.02) = -0.040	p = 0.968	ns
Ext. Data Figure 5	B	**Typ ITI** (n = 19) **Typ ITI-few** (n = 12)	change in lick rate to cue, trials 36-40	Welch's t-test	t (28.59) = 0.89	p = 0.762 (Bonferroni corrected)	ns
		Ext ITI (n = 19) **Typ ITI-few** (n = 12)	change in lick rate to cue, trials 36-40	Welch's t-test	t (28.55) = -7.86	$p = 2.55 \times 10^{-08}$ (Bonferroni corrected)	****

Ext. Data Figure 7	C	**Typ ITI** (n = 5) **Ext ITI** (n = 7)	DA learned trial	Welch's t-test	t (4.02) = 4.46	p = 0.0110	*
Ext. Data Figure 9	A	**Typ ITI** (n = 20) **Ext ITI** (n = 20)	simulation - norm. lick to cue	Welch's t-test	t (31.21) = -44.15	$p = 1.06 \times 10^{-29}$	****
Ext. Data Figure 9	B	**Typ ITI** (n = 20) **Ext ITI** (n = 20)	simulation - DA learned trial	Welch's t-test	t (20.06) = 80.46	$p = 1.17 \times 10^{-26}$	****