



Bridging Computation and Representation in Associative Learning

Samuel J. Gershman¹

Accepted: 26 March 2025 / Published online: 4 April 2025
© Society for Mathematical Psychology 2025

Abstract

Two views of Pavlovian conditioning have dominated the theoretical discourse. The classical associative view holds that associations are learned based on temporal contiguity between stimuli, and conditioned responses directly reflect associative strength. The representational view, exemplified by Rate Estimation Theory (Gallistel & Gibbon, *Psychological Review*, 107(2), 289–344 2000), holds that animals learn the structure of the stimulus distribution, from which a measure of contingency between stimuli is derived and used to generate conditioned responses. Unlike contiguity, contingency is a relative measure, comparing the rate of reinforcement in the presence of a stimulus to the background rate. This turns out to be crucial for explaining the effects of manipulating the background rate while holding the stimulus-conditional rate constant (i.e., changing contingency without changing contiguity). It has also been argued that contiguity theories face irremediable conceptual difficulties stemming from the coercion of continuous time into discrete bins. This paper makes two contributions to the debate. First, it shows that Rate Estimation Theory faces its own computational and conceptual problems. Second, it shows how to fix these problems while retaining the core of the theory. Surprisingly, this leads to the insight that rates can be estimated using an algorithm closely resembling a classical associative theory (the Rescorla-Wagner model). The key difference lies in the response rule rather than in the learning rule.

Keywords Associative learning · Pavlovian conditioning · Bayesian inference

Introduction

The concept of association has played a central role in both the psychology and neurobiology of learning, particularly as applied to Pavlovian conditioning. In a typical Pavlovian delay conditioning protocol, a conditioned stimulus (CS, such as a tone) is presented for some duration, after which an unconditioned stimulus (US, such as a food pellet) is delivered. With repeated pairings, the animal comes to produce a conditioned response (CR) to the CS (e.g., anticipatory head entries into the feeder trough). This is commonly considered a paradigmatic example of associative learning—a theory-laden descriptor of both the experimental protocol and the underlying psychological/neurobiological process. It implies that the animal produces a CR because it has formed an association between the CS and the US, realized neurally via synaptic plasticity.

The idea that associations underlie conditioned responding is both deeply entrenched and deeply problematic, as reviewed in the next section. An important alternative to the associative view, championed by Gallistel and his colleagues (Gallistel, 1990; Gallistel & Gibbon, 2000), is a representational view of learning, according to which animals acquire and use facts about the structure of the stimulus distribution. In the context of Pavlovian conditioning, this view posits that animals estimate conditional rates (i.e., the number of USs per unit of continuous time contributed by a particular stimulus). In the model of Gallistel and Gibbon (2000), a CR is generated when the CS rate exceeds the background rate by some threshold. More recent work has conceptualized the decision process in information-theoretic terms, where the mutual information between the CS and the US formalizes a notion of temporal contingency (Balsam et al., 2006; Balsam & Gallistel, 2009; Ward et al., 2012; Gallistel et al., 2014, 2019; Gallistel & Latham, 2023; Kalmbach et al., 2019). These models have successfully explained a wealth of quantitative data on conditioning (but see Bouton & Sunsay, 2003; Harris et al., 2019; Austen & Sanderson, 2020, for some empirical challenges). They have also overcome some

✉ Samuel J. Gershman
gershman@fas.harvard.edu

¹ Department of Psychology and Center for Brain Science,
Harvard University, Cambridge, USA

of the conceptual challenges facing both association-based models and prior definitions of contingency.

Despite their success, there are several important theoretical issues that need to be settled. One issue is how to estimate the rates in a way that is both computationally plausible and mathematically defensible. A second issue is how to more rigorously relate the quantitative empirical laws of conditioning to the predictions of the model. In this paper, we address these two issues in a unified way, starting with a new analysis of the rate estimation problem that yields a simple error-driven learning algorithm. We then show how the resulting rate estimates, when combined with a decision rule, give rise to an important quantitative empirical law: the timescale invariance of learning, as explained in the next section. Finally, we quantitatively assess the model predictions on several datasets.

From Contiguity to Contingency

In most associative theories of learning, temporal contiguity between stimuli is a necessary condition for the formation of an association. For this to be well-defined, we need to specify what counts as temporally contiguous—what is the critical interstimulus interval? At first glance, there does seem to be a critical interval, which differs across experimental protocols (see Rescorla, 1988). For a given experimental protocol (see Fig. 1 for an illustration), there is a particular CS-US interval which produces the fastest CR acquisition, where acquisition speed is typically measured as the inverse number of reinforcements until the CR rate meets an acquisition criterion. It is important to note, however, that this assumes a fixed intertrial interval. If the intertrial interval is rescaled to maintain a fixed ratio with the interstimulus interval, then the acquisition speed is constant across different interstimulus intervals, a phenomenon known as *timescale invariance* (Gibbon et al., 1977; Gallistel & Gibbon, 2000).¹ Put another way, *there is no critical interstimulus interval*. This fact seems devastating for the claim that temporal contiguity is a necessary condition for association formation and hence conditioned responding.

One possible remedy is to invoke cue competition. In addition to the CS, we can posit a constant “background” stimulus which also forms an association with the US. The CR is determined by the combination of CS and background associations. The CS and background stimulus also compete with one another during learning, such that credit for the US is split between the CS and background stimulus. The intertrial interval can then be viewed as an extinction period for the

background stimulus, weakening its association and thereby allowing the CS to take more credit for the US. Unfortunately, this won’t (without further assumptions) solve the problem of timescale invariance; the CS association is strengthened and the background association is weakened, but these two changes cancel each other out in the CR, which combines the CS and background associative strengths.

The Rescorla-Wagner Model

To make the preceding point concrete, let’s look at the most influential model of associative learning, the Rescorla-Wagner model (Rescorla & Wagner, 1972). This model operates in discrete time, so we need to divide continuous time into bins of some size, which we index by n (we will return to the issue of discrete vs. continuous time shortly). We use $r(n) \in \{0, 1\}$ to denote the US delivered at time n , and $\hat{r}(n)$ to denote the model’s prediction of this US. The prediction is a linear combination of stimulus associative strengths $w_i(n)$ and the stimulus presence $x_i(n) \in \{0, 1\}$ for stimulus i :

$$\hat{r}(n) = \sum_i w_i(n)x_i(n). \quad (1)$$

We include in the set of stimuli a “background” stimulus B which is constantly present; implicitly, $x_B(n) = 1$ for all n . The CR is typically assumed to be a monotonic function of the US prediction,² though it should be noted that non-associative factors (e.g., habituation) can also affect the CR (see, for example, Reiss & Wagner, 1972). Once $r(n)$ is observed, the associative strength is updated based on the prediction error signal $\delta(n) = r(n) - \hat{r}(n)$:

$$w_i(n+1) = w_i(n) + \alpha x_i(n)\delta(n), \quad (2)$$

where $\alpha \in [0, 1]$ is a learning rate.³ We will study this model in the simple setting where there is a single cue (which we denote by i) along with the background stimulus b . Using this simple model, we examine the effects of changing the interstimulus and intertrial intervals on the US prediction (and by extension the CR). If we define the acquisition criterion as some fraction of the asymptotic US prediction, $\hat{r}(\infty)$,

¹ Some studies have shown that timescale invariance breaks under some circumstances (Lattal, 1999; Holland, 2000) but as pointed out by Ward et al. (2012), these results should be interpreted as caution, because their dependent variable was not reinforcements to acquisition.

² Rescorla and Wagner preferred to remain agnostic about the mapping from associations to behavior. However, as discussed in Miller et al. (1995), monotonicity has been the operational assumption in nearly all applications of the model. More recent work has investigated a variety of monotonic response functions (Thein et al., 2008; Holmes et al., 2019; Ghirlanda, 2022).

³ In their original work, Rescorla and Wagner presented a slightly more complex model, where the learning rate is decomposed into stimulus-specific and reinforcement-specific components. We won’t discuss these complexities, except briefly when addressing the relative influence of context and CS on learning.

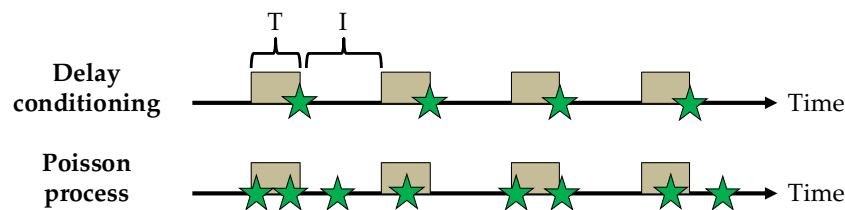


Fig. 1 Conditioning protocols. (Top) standard Pavlovian delay conditioning. (Bottom) conditioning where reinforcements are generated according to a Poisson process. Boxes represent periods during which the conditioned stimulus (CS) is present. Stars represent instanta-

neous unconditioned stimulus (US) delivery. The interstimulus interval between CS onset and US delivery is denoted by T , and the intertrial interval between US delivery and the next CS onset is denoted by I . The cycle time C is the sum of these two intervals

then acquisition speed will be a monotonic function of $\hat{r}(\infty)$. We can thus focus our analysis on the asymptotics of the Rescorla-Wagner model.

For the single cue setting, the asymptotic weights for the background and CS can be found analytically (Chapman & Robbins, 1990; Gallistel, 1990; Danks, 2003):

$$w_B(\infty) = \frac{\bar{R} - \bar{R}_{CS}}{C - T} \quad (3)$$

$$w_{CS}(\infty) = \frac{\bar{R}_{CS}}{T} - w_B(\infty), \quad (4)$$

where T is the interstimulus (CS-US) interval, C is the US-US interval (the sum of the interstimulus and intertrial intervals, also known as the *cycle time*), \bar{R}_{CS} is the average number of reinforcements per trial during the CS, and \bar{R} is the average number of total reinforcements per trial (including the intertrial interval). Technically, the solution depends on the time discretization, where each timestep n increments time by Δt ; since the discretization only changes the US prediction by a scale factor that doesn't depend on any of the experimental parameters, we implicitly set it to 1. However, time discretization will come back to bite later.

In a standard delay conditioning protocol, $\bar{R}_{CS} = \bar{R} = 1$, so that $w_{CS}(\infty) = 1/T$ and $w_B(\infty) = 0$. This conforms to the intuition that the CS should receive all of the credit since the US only appears during the CS. Critically, $\hat{r}(\infty) = w_{CS}(\infty) + w_B(\infty) = 1/T$, which means that the asymptotic US prediction is *not* timescale invariant.

More generally, $\hat{r}(\infty) = \bar{R}_{CS}/T$ for the single CS setting. In other words, the asymptotic US prediction will always report the CS-conditional reinforcement rate. This means that the Rescorla-Wagner model will be insensitive to any manipulation of CS-US contingency that leaves the CS-conditional reinforcement rate intact. Several experimental protocols have been used to show that in fact the CR can be increased or decreased by changes in the background reinforcement rate while holding the CS-conditional rate constant (see Escobar & Miller, 2004, for a review). For example, adding unsignaled US deliveries during the

intertrial interval (Rescorla, 1968; Gamzu & Williams, 1971; Lindblom & Jenkins, 1981), prior to conditioning (Randich & LoLordo, 1979; Balsam & Schwartz, 1981; Overmier et al., 1979), or after conditioning (Rescorla, 1973; Overmier et al., 1979), all have the effect of decreasing the CR rate. It can also be increased by lengthening the intertrial interval (i.e., spacing; Terrace et al., 1975; Gibbon et al., 1977; Sunsay & Bouton, 2008), pre-exposing an animal to the experimental context (Lattal & Abel, 2001), or extinguishing the context following conditioning (Yin et al., 1993; Miguez et al., 2011).

At first glance, it might seem that even if the asymptotic US prediction is invariant to these manipulations, the *speed* of learning should be affected. Intuitively, increasing the ITI should effectively act as an extinction phase for the background, thereby increasing the positive prediction error when the US is delivered during the CS. While this is true, the background associative strength is *also* reinforced by the same prediction error, and this compensates for the extinction during the ITI. As shown in the left panel of Fig. 2, doubling the ITI has virtually no effect on the learning curve. The same line of reasoning explains why contingency degradation (US delivery during the ITI) has no effect on the learning curve (Fig. 2, right panel).

In these simulations, we assumed equal learning rates for the context and CS. However, Rescorla and Wagner (1972) assumed that the learning rate for context is smaller. Figure 3 shows that this does not solve the problem. The learning curves for different ITIs are still perfectly overlapping, and the learning curves for different levels of contingency are slightly separated *in the wrong direction* (i.e., faster learning in the degraded contingency condition).

A knowledgeable reader might at this point object: didn't Rescorla and Wagner successfully simulate contingency degradation in their original work? Yes and no. They did indeed simulate contingency degradation, but they only showed what happens to the associative strength of the CS, not the summed associative strengths. This would be fine if the response model only took into account the CS associative strength. However, doing so leads to other problems. For example, Rescorla (2000) showed that CS and context

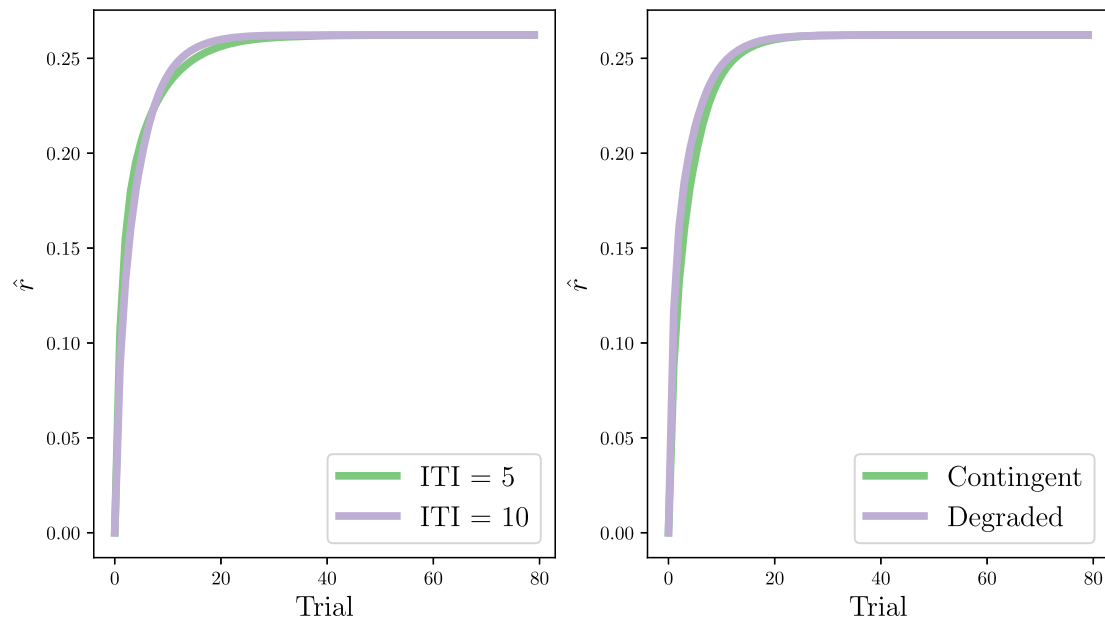


Fig. 2 The Rescorla-Wagner model cannot explain spacing and contingency degradation effects. Each plot shows the US prediction \hat{r} as a function of the training trial. The learning rates for context and CS were both set to 0.1. (Left) Learning curves are invariant to changing the

intertrial interval (ITI). (Right) Learning curves are invariant to delivering a US during the ITI (contingency degradation). The “Contingent” curve shows results for the standard delay conditioning protocol with reinforcements delivered only at CS offset

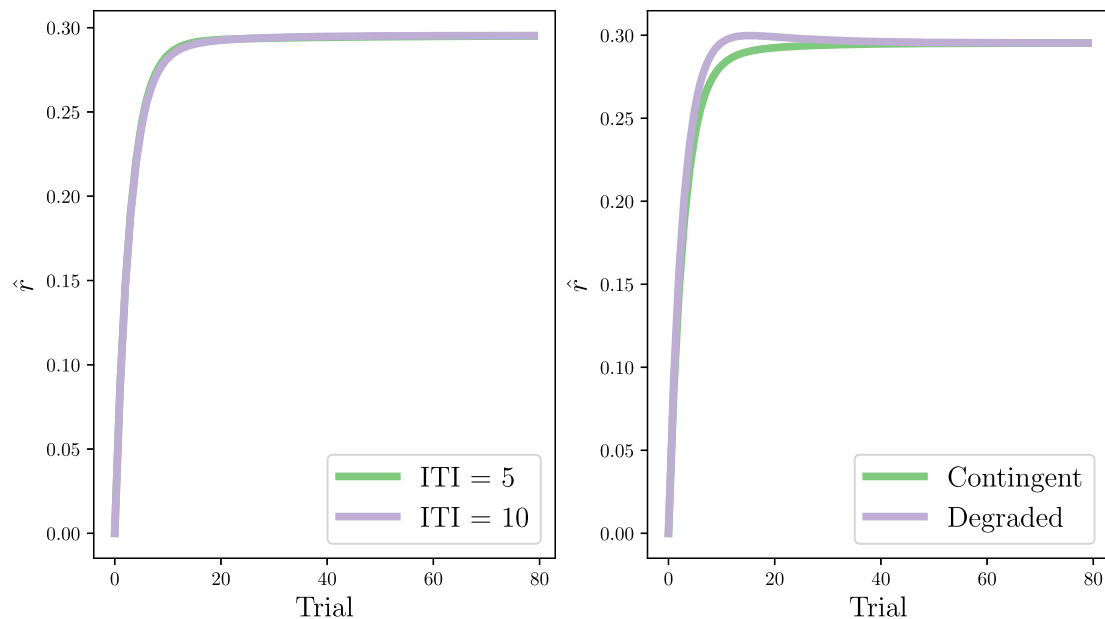


Fig. 3 The Rescorla-Wagner model cannot explain spacing and contingency degradation effects, even with different learning rates for the context and CS. Each plot shows the US prediction \hat{r} as a function of the training trial. The learning rate for the context was set to 0.01 and the learning rate for the CS was set to 0.1. (Left) Learning curves are invari-

ant to changing the intertrial interval (ITI). (Right) Learning curves are invariant to delivering a US during the ITI (contingency degradation). The “Contingent” curve shows results for the standard delay conditioning protocol with reinforcements delivered only at CS offset

responding were similar after training with a random control procedure, where the US delivery rate was the same in the presence and absence of the CS, recapitulating prior results (Rescorla, 1968). Critically, testing the CS in a novel context produced a higher rate of responding to the CS compared to context alone. Since the CS is the same in both contexts, a model which bases responding only on the CS associative strength would have no means of explaining such context-dependency. At a more conceptual level, such a model raises questions about how to explain context responding; does context only drive responding in the absence of the CS but not in the presence of the CS? Answering this question would seem to require a more complex response rule.

All of these observations suggest that temporal contiguity—the degree to which the CS and US co-occur in time—is inadequate as a principle of Pavlovian *responding* (i.e., the conditioned response to a CS cannot be predicted by temporal contiguity alone). On the other hand, it is unclear whether temporal contiguity is inadequate as a principle of Pavlovian *learning* (i.e., contiguity may be sufficient to predict what information animals extract from their experience). To appreciate the difference between these two claims, consider the class of models that generate CRs based on the relative strength of the CS-US relationship and a *comparator*, such as the background-US relationship (Miller & Escobar, 2001). Rate Estimation Theory (RET; reviewed in the next section) falls into this class, as does the sometimes-competing retrieval model (Stout & Miller, 2007; Ghirlanda & Ibadullayev, 2015) and Rescorla’s semi-formal contingency model (Rescorla, 1967). The stimulus relationships are variously interpreted by these models as rates, associations, or probabilities; the next section discusses the relationships between these quantities. We have already seen that the Rescorla-Wagner model, which is quintessentially associative, can be interpreted as estimating the CS-conditional reinforcement rate up to a constant determined by the discretization of time, which is identical to the CS-conditional reinforcement probability. Regardless of how the stimulus relationships are interpreted or how they are learned, the critical feature of these models is their relative response rule, which stipulates that some form of contingency generates conditioned responding. Intuitively, if the US rate during the CS matches its rate during the background, then the CS tells the animal nothing new about the US rate, even if the US reliably occurs in the presence of the CS. If the US rate increases upon the appearance of the CS, then it is natural to say that the US is contingent on the CS. These intuitions are captured by relative response rules.

What Is Contingency?

What exactly does contingency mean, and are different models talking about the same thing when they use this term? In

modal logic, a proposition is designated as contingent if it is possible but not necessarily true; there must be at least one “possible world” in which the proposition is true. We then say that event A is contingent on event B if B occurs in all the possible worlds in which A occurs. This definition is closely related to counterfactual theories of causation (Lewis, 1973), according to which B causally depends on A if and only if: (i) if B were to occur, then A would occur (sufficiency), and (ii) if B were to not occur, then A would not occur (necessity). Pearl (2000) developed a probabilistic theory of counterfactual dependence for random variables. Probabilistic counterfactuals have played an important role in modern psychological theories of causal judgment (Gerstenberg, 2024).

We can connect these ideas to animal learning theory by showing (under some assumptions) that the Rescorla-Wagner model estimates a particular probabilistic counterfactual, the probability that the CS is both necessary and sufficient for producing the US. In particular, if the CS is excitatory (it never reduces the probability of the US) and there are no hidden confounders (i.e., latent variables that might cause both the CS and the US),⁴ then the probability of necessity and sufficiency is given by:

$$\Delta P = P(r = 1|x_{CS} = 1) - P(r = 1|x_{CS} = 0), \quad (5)$$

where we have dropped the timestep index for the CS. We use the notation ΔP to draw attention to the fact that this equation is identical to the definition of contingency used extensively in the literature on human causal judgment (Jenkins, 1965), and has also been used to analyze Pavlovian conditioning (Gibbon et al., 1974). Inspection of Eq. 4 reveals that the asymptotic cue weight learned by the Rescorla-Wagner model in the single cue setting is $\Delta P = w_{CS}(\infty)$, a fact also noted by Chapman and Robbins (1990). This further implies that ΔP can be computed by contrasting the US prediction in the presence and absence of the US—i.e., a relative response rule. We have thus come full circle, linking the Rescorla-Wagner model to a rigorous causal definition of contingency that coincides with a particular relative response rule.

It is important to note that ΔP is not timescale invariant: multiplying both T and C by a constant will not yield the same value of ΔP . There is a related problem (Gallistel, 2021): Recall that using probabilities to define contingency for Pavlovian conditioning requires us to pick a time discretization, Δt . This seems rather innocuous, but it isn’t. If Δt becomes infinitesimally small, we get the following

⁴ See Pearl (2000), Chapter 9, for details. Pearl refers to the excitatory requirement as *monotonicity* and the no confounding requirement as *exogeneity*.

expression for ΔP :

$$\lim_{\Delta t \rightarrow 0} \frac{\bar{R}_{CS}}{T/\Delta t} - \frac{\bar{R} - \bar{R}_{CS}}{(C - T)/\Delta t} = 0, \quad (6)$$

where division by Δt translates continuous time into a number of discrete timesteps. Thus, the definition implies that the strength of the CR will be independent of the experimental parameters, and in fact, no conditioned responding should occur at all in this limit. Alternatively, one could make the discretization coarser, but then one runs the risk of having more than one US occur in a single time bin, violating the assumption (implicit in the above treatment) that the distributions are defined over binary events.

A tempting solution to both these problems is to use the ratio between conditional probabilities rather than the difference. This eliminates the time discretization factor, and it satisfies timescale invariance. However, recall that $w_B(\infty)$ is 0 for the delay conditioning protocol. This means that the ratio is ill-defined.

The nub of the problem, as pointed out repeatedly by Gallistel, is that Pavlovian conditioning (like many naturalistic learning events) occurs in continuous time. Shoe-horning it into a discrete-time probability distribution has disastrous consequences. A better alternative, considered next, is a continuous-time treatment.

Rate Estimation Theory

The critical conceptual step undertaken by RET (Gallistel, 1990; Gallistel & Gibbon, 2000) is the replacement of discrete timesteps (n) with continuous time (t). Some events, such as US deliveries (reinforcements), are well-characterized as point processes: they occur near-instantaneously (and for our purposes, we model them as instantaneous) at particular points in time. Other events, such as a typical CS, are interval events: they endure for some period of time. RET frames the computational problem facing an animal as one of predicting the US patterns from the CS patterns and using these predictions to decide when to respond. RET makes a set of structured assumptions about the CS-US relationship which reduces the problem to estimation of CS-conditional US rates. In the following subsections, we describe the problem and assumptions, an algorithmic solution proposed by Gallistel (1990) for solving the problem, and how it addresses the empirical issues raised earlier. This will lead us into a discussion of RET's limitations—the starting point for a different algorithmic solution.

Problem Statement

We consider an animal that observes a point process of reinforcements, $r(t)$, where t indexes time (illustrated in Fig. 1). We impute to the animal the following generative model:

- The reinforcement process can be additively decomposed into component processes, each of which is excited by the presence of a stimulus:

$$r(t) = \sum_i r_i(t), \quad (7)$$

where i indexes stimuli.

- Each component r_i is generated by a Poisson process with intensity

$$\bar{r}_i(t) = \mathbb{E}[r_i(t)|x_i(t)] = \lambda_i x_i(t), \quad (8)$$

where λ_i is the reinforcement rate in the presence of stimulus i , and $x_i(t) = 1$ when the stimulus is present (0 otherwise).⁵ This implies that $r_i(t) = 0$ whenever $x_i(t) = 0$.

These two assumptions together imply that $r(t)$ follows a Poisson process with intensity

$$\bar{r}(t) = \mathbb{E}[r(t)|x(t)] = \sum_i \lambda_i x_i(t). \quad (9)$$

The animal observes the stimulus process $\mathbf{x}(t)$ and the reinforcement process $r(t)$, but not the underlying components. The learning problem facing the animal is to estimate the rates for each component.

An Algorithmic Solution

Gallistel (1990) devised a clever algorithmic solution to the rate estimation problem. The key idea is to take advantage of rate additivity (Eq. 9), turning rate estimation into a linear system identification problem. Specifically, rate additivity implies the following relation:

$$\mathbb{E} \left[\frac{\tilde{R}_i(t)}{N_i(t)} \right] = \sum_j \lambda_j \mathbb{E} \left[\frac{N_{ij}(t)}{N_i(t)} \right], \quad (10)$$

where $\tilde{R}_i(t) = \int_{\tau=0}^t x_i(\tau) r(\tau) d\tau$ is the total number of reinforcements observed in the presence of stimulus i , $N_i(t) =$

⁵ More generally, stimuli are allowed to have continuous intensities. However, we will not need to make use of continuous intensities in this paper.

$\int_{\tau=0}^t x_i(\tau) d\tau$ is the cumulative record of stimulus i , and $N_{ij}(t) = \int_{\tau=0}^t x_i(\tau)x_j(\tau) d\tau$ is the cumulative pairwise record of stimulus i and j (i.e., the total amount of time during which they were presented together).

Gallistel refers to the quantity $u_i(t) = R_i(t)/N_i(t)$ as the “uncorrected” rate estimate for stimulus i . Its expectation equals the true rate λ_i only when stimulus i is consistently presented alone, so that $N_{ij} = 0$ for $i \neq j$ (which essentially never happens since we assume a constant background stimulus). The uncorrected rate estimate needs to be corrected for the influence of other stimuli on the observed reinforcement process. Since both the uncorrected rate estimate and the cumulative pairwise records are observable by the animal, standard linear algebra can be used to estimate the corrected rates:

$$\hat{\lambda}(t) = \mathbf{A}(t)^{-1} \cdot \mathbf{u}(t), \quad (11)$$

where $\hat{\lambda}$ is a column vector containing the corrected rate estimates, $\mathbf{u}(t)$ is a column vector containing the uncorrected rate estimates, and $\mathbf{A}(t)$ is a matrix containing the normalized cumulative pairwise records, $A_{ij}(t) = N_{ij}(t)/N_i(t)$.

While mathematically sound, there are several drawbacks to this approach. First, it requires tracking co-occurrence statistics for all stimuli, a memory demand (space complexity in the jargon of computer science) that is quadratic in the number of stimuli. Second, it requires tracking these statistics over a long enough time period that \mathbf{A} and \mathbf{u} match their expectations. Third, the algorithm has no way of tracking uncertainty in the estimates after a finite observation period, which is needed to determine the reliability of future predictions. Finally, the linear system needs to be solved repeatedly, which has a time complexity that is between quadratic and cubic in the number of stimuli, depending on the implementation. What’s needed is an algorithm that can operate in real time, with space and time complexity that doesn’t scale super-linearly with the number of stimuli. Ideally, the algorithm should also track estimation uncertainty.

The Decision Rule

To produce conditioned behavior, RET assumes that the rate estimates $\hat{\lambda}$ are translated into conditioned responses based on a comparison of the CS and background rate estimates. Specifically, Gallistel and Gibbon (2000) proposed that an animal responds to the CS whenever

$$\frac{\hat{\lambda}_{\text{CS}}(t) + \hat{\lambda}_{\text{B}}(t)}{\hat{\lambda}_{\text{B}}(t)} > \beta, \quad (12)$$

where β is a threshold parameter.

Equation 12 can be given an information-theoretic interpretation. Under the assumptions of the model, the distribution

over the time to the next reinforcement, d , is an exponential distribution with mean $1/(\lambda_{\text{CS}} + \lambda_{\text{B}})$ if the CS is present, and $1/\lambda_{\text{B}}$ if the CS is absent. We can express the model’s uncertainty about d (in the absence of the CS) using the differential entropy of the exponential distribution:

$$H[d|x_{\text{CS}} = 0] = 1 + \log(\lambda_{\text{B}}). \quad (13)$$

The differential entropy conditional on the CS is defined similarly:

$$H[d|x_{\text{CS}} = 1] = 1 + \log(\lambda_{\text{CS}} + \lambda_{\text{B}}). \quad (14)$$

The difference between Eqs. 13 and 14 expresses the “information gain” due to observing the CS (Lindley, 1956)—i.e., the degree to which uncertainty about the time to next reinforcement is reduced by the appearance of the CS:

$$\Delta H = H[d|x_{\text{CS}} = 1] - H[d|x_{\text{CS}} = 0] = \log \frac{\lambda_{\text{CS}} + \lambda_{\text{B}}}{\lambda_{\text{B}}}. \quad (15)$$

We can see from this expression that it is equivalent to the log transformation of the decision variable in Eq. 12, substituting the rate estimates for the true rates. This derivation is similar to the one given by Balsam et al. (2006), except that they use the entropy of the marginal distribution $H[d]$ instead of the conditional distribution $H[d|x_{\text{CS}} = 0]$. This entails separately estimating the US rate while ignoring the CS completely. They refer to this as the *contextual rate of reinforcement*; this is equivalent to the background rate of reinforcement, λ_{B} , for a model in which only the background stimulus occurs.⁶ In the treatment given here, only a single model is learned, rather than separate models with and without the CS.

Equation 15 is closely related to Bayesian updating; it can be understood as the Kullback–Leibler divergence (relative entropy) between the posterior distribution (after the appearance of the CS) and prior distribution (before the appearance of the CS) over the time to next reinforcement. The expectation of Eq. 15 with respect to the CS distribution, $p(x_{\text{CS}})$, yields the mutual information between x_{CS} and d . Normalizing Eq. 15 by the entropy yields the measure of contingency proposed by Gallistel and Latham (2023).

Note that the information gain defined above reflects uncertainty only about the time to the next reinforcement given a set of rates; it does not reflect uncertainty about

⁶ The contextual rate of reinforcement corresponds to the uncorrected rate estimate for the background stimulus. So in the linear system approach to rate estimation, it is already computed as an intermediate quantity.

unknown rates. Recently, Gallistel and Latham (2023) have formulated a decision variable that takes into account rate uncertainty (more on this later).

The relative response rule formalized by information gain plays an important role in explaining the empirical phenomena that are so problematic for classical models of associative learning like the Rescorla-Wagner model, which assumes that conditioned responding directly reflects associative strength. As discussed above, manipulations like contingency degradation and US preexposure reduce conditioned responding without affecting the CS-US contiguity (and hence associative strength). These phenomena, which increase $\hat{\lambda}_B(t)$, are naturally explained by the relative response rule. It is also at the heart of how RET explains timescale invariance.

Explaining Timescale Invariance

Gallistel and Gibbon (2000) derived timescale invariance from RET as follows. Recall that we are considering the standard Pavlovian delay conditioning setting, where the US always follows the CS after a fixed delay. Because of rate additivity, we have $\lambda_{CS} + \lambda_B = 1/T$, where T is the CS-US interval. This follows from the fact that in delay conditioning exactly 1 reinforcement occurs during the CS-US interval. Gallistel and Gibbon further assumed that $\lambda_B = 1/(RI)$, where I is the intertrial interval and R is the number of trials (equivalent here to the number of reinforcements). In other words, the background reinforcement rate is assumed to be the reciprocal of the total background exposure alone. Note that this assumption is not derived from RET—it is essentially *ad hoc*. With this assumption in place, along with the assumption that the rate estimates have accurately recovered the true estimates ($\hat{\lambda} = \lambda$), RET predicts that a conditioned response will be produced when

$$\frac{RI}{T} > \beta. \quad (16)$$

Rearranging, the number of reinforcements to acquisition (i.e., the transition point between not responding and responding) is given by:

$$R^* = \beta \left(\frac{I}{T} \right)^{-1}. \quad (17)$$

We thus recover the important empirical law that trials (reinforcements) to acquisition (R^*) is linearly related to the ratio of interstimulus and intertrial intervals on a logarithmic scale, such that rescaling both intervals by a constant leaves the acquisition speed unchanged (Gibbon et al., 1977; Gibbon & Balsam, 1981; Gallistel & Gibbon, 2000; Harris & Gallistel, 2024). Note that in some formulations of

timescale invariance, the law is a function of C/T , a protocol's *informativeness* (Balsam & Gallistel, 2009), rather than I/T . For example, Harris and Gallistel (2024) describe the law as $R^* = \beta(\frac{C}{T} - 1)^{-1}$. Because $\frac{C}{T} - 1 = \frac{I}{T}$, these laws are equivalent. We will revisit the functional form of timescale invariance in a later section.

While the derivation of timescale invariance from RET is an elegant and satisfying result, we need to acknowledge two limitations. First, as already mentioned, it relies on an *ad hoc* assumption about the background rate. Second, it does not take into account the animal's uncertainty about the rates, and in fact, assumes that the rates have been perfectly estimated. Gallistel and Latham (2023) developed one method for dealing with this issue, but this relies on other *ad hoc* assumptions. In particular, their decision rule is not derived from first principles, but designed based on some intuitive desiderata. Furthermore, they do not present an analytical derivation of timescale invariance from their decision rule. Thus, the problem appears to remain unsolved.

Summary

This section has reviewed how RET addresses some of the fundamental problems with classical associative learning theory. First, it avoids the pathologies of discrete time by formulating the learning problem in continuous time. Second, it explains a wide range of challenging findings by using a relative response rule, which compares the CS-conditional and background rates of reinforcement. The information-theoretic version of this model provides a formal definition of temporal contingency.

Several general issues vex RET in its existing forms. First, the algorithm for estimating rates is computationally impractical. Second, the derivation of timescale invariance involves some questionable assumptions and does not take into account the animal's uncertainty about rates. We next turn to a new approach which addresses these issues.

A New Approach to the Rate Estimation Problem

Our goal in this section is to propose a new version of RET which retains its essential idea (that animals are estimating conditional rates of reinforcement) but replaces its algorithmic machinery. In doing so, we remedy some of the shortcomings of the theory from a mathematical and computational perspective. The new approach also allows us to draw a connection between rate estimation and error-driven learning models. Finally, we show analytically how an information gain decision rule gives rise to timescale invariance.

Maximum Likelihood Estimation

Let's start with a simpler problem: rate estimation when the components are observed. We will shortly see how to finesse the credit assignment problem with multiple hidden components into approximately this form.

Suppose at time t the animal has an estimate $\hat{\lambda}_i(t)$. The maximum likelihood estimate of the rate can be obtained in continuous time using a closed-form, recursive update:

$$\frac{d\hat{\lambda}_i(t)}{dt} = \frac{x_i(t)[r_i(t) - \hat{\lambda}_i(t)]}{N_i(t)}, \quad (18)$$

where, as above, $N_i(t)$ is the cumulative record of stimulus i . The update equation can be derived by first expressing the maximum likelihood estimator recursively at two update times (t_0 followed by t):

$$\hat{\lambda}_i(t) = \frac{R_i(t)}{N_i(t)} = \hat{\lambda}_i(t_0) + \frac{R_i(t) - R_i(t_0) - \hat{\lambda}_i(t_0)}{N_i(t)}, \quad (19)$$

where $R_i(t) = \int_0^t x_i(\tau)r_i(\tau)d\tau$ is the cumulative record of reinforcements for component i . Taking the continuous-time limit $t - t_0 \rightarrow 0$, we obtain Eq. 18.

Estimation when the Components Are Unobserved

Let's now return to the original credit assignment problem. Because of rate additivity, we can use our rate estimates to obtain an estimate of the components:

$$r_i(t) \approx \delta(t) + \hat{\lambda}_i(t)x_i(t), \quad (20)$$

where

$$\delta(t) = r(t) - \sum_j \hat{\lambda}_j(t)x_j(t) \quad (21)$$

is the *global prediction error*. This is a stochastic approximation of the following equality that holds when $\hat{\lambda} = \lambda$:

$$\mathbb{E}[r_i(t)] = \mathbb{E}[\delta(t)] + \lambda_i x_i(t), \quad (22)$$

where $\mathbb{E}[\delta(t)] = 0$ and the expectation is taken over realizations of the Poisson process. Plugging the approximation into Eq. 18 yields:

$$\begin{aligned} \frac{d\hat{\lambda}_i(t)}{dt} &\approx \frac{x_i(t)[\delta(t) + \hat{\lambda}_i(t)x_i(t) - \hat{\lambda}_i(t)]}{N_i(t)} \\ &= \frac{x_i(t)\delta(t)}{N_i(t)}, \end{aligned} \quad (23)$$

where we have used the assumption that $x_i(t) \in \{0, 1\}$, which means that $\hat{\lambda}_i(t)x_i(t) - \hat{\lambda}_i(t) = 0$ when $x_i(t) = 1$. Equation 23 is remarkably similar to the Rescorla-Wagner update, but defined in continuous time, with learning rate $1/N_i(t)$. Rather than updating associative strengths, it updates rate estimates.⁷

There are several ambiguities in this algorithm that need to be resolved. First, it leaves open how the rate estimates are initialized. It turns out that this ambiguity can be normatively resolved by looking at the problem through a Bayesian lens. Another ambiguity is that $\delta(t)$ is a function of the entire stimulus history, not just a single stimulus, and consequently, it is not clear whether the learning rate should be $1/N_i(t)$, which is taken from the fully observable case where we could track the reinforcement records separately for each stimulus. We will try to resolve this issue in the next section. Finally, there is an ambiguity about when updating should happen; presumably, it is happening intermittently, but we lack strong empirical constraints on exactly which events induce updating. In the simulations reported below, it is assumed (for convenience) that updating happens at regular 500 ms intervals.

Bayesian Estimation

Eq. 23 can, with little modification, accommodate Bayesian estimation. The standard Bayesian analysis for Poisson processes (see Gelman et al., 2013) uses a Gamma distribution over rates, which is the conjugate prior (so that the posterior is also a Gamma distribution). Formally, let $\lambda_i \sim \text{Gamma}(r_0, n_0)$ with “shape” parameter $r_0 > 0$ and “inverse scale” parameter $n_0 > 0$. The notation was chosen to make transparent a particular interpretation of these parameters: the shape parameter can be interpreted as the effective number of prior reinforcements, and the inverse scale parameter can be interpreted as the effective observation period. Accordingly, the rates are initialized as $\hat{\lambda}_i(0) = r_0/n_0$ and n_0 is added to the cumulative stimulus record so that the update is given by:

$$\frac{d\hat{\lambda}_i(t)}{dt} = \frac{x_i(t)\delta(t)}{N'_i(t)}, \quad (24)$$

where $N'_i(t) = N_i(t) + n_0$.

We now need to address the ambiguity (identified in the previous section) of how to set the learning rate. At one extreme, we can use $1/N'_i(t)$, as we've done in Eq. 24, which assumes independent learning rates for each stimulus based on its individual stimulus record. The other extreme would be to use $1/t$, which assumes a shared maximal observa-

⁷ Another way to arrive at a similar result is to assume a squared error loss function and then update the rate estimates using stochastic gradient descent.

tion period for all stimuli. As a compromise, we will adopt the following formula: $N'_i(t) = \eta_i t + n_0$. This is a shared learning rate for all stimuli, but decremented by a parameter $\eta_i \in [N_i(t)/t, 1]$ to reflect the fact that the effective observation period is somewhere in between the fully independent and fully shared cases. In practice, we fix η_i to a constant value across stimuli.

The rate estimates can be understood as the posterior means (or more precisely an approximation of the posterior means, when using the results from the preceding section). The Bayesian estimates approximate maximum likelihood estimates in the limit $r_0 \rightarrow 0, n_0 \rightarrow 0$. Importantly, the Bayesian setup resolves the ambiguity about initial conditions left open by maximum likelihood estimation.

One consequence of Bayesian estimation is sensitivity to sample size (or duration). Intuitively, an animal should be more confident if it has observed the process for longer. In contrast, maximum likelihood estimation will yield the same rate estimates for the same empirical rates $R_i(t)/N_i(t)$ regardless of how long the process has been observed. To see this, we represent the Bayesian estimate in the following form:

$$\hat{\lambda}_i(t) = \omega \frac{R_i(t)}{N_i(t)} + (1 - \omega) \frac{r_0}{n_0}, \quad (25)$$

where

$$\omega = 1 - \frac{n_0}{N'_i(t)} \quad (26)$$

is the weight on the empirical rates. Because the posterior mean rate is a convex combination of the empirical and prior rates, weighted inversely by sample size, it will be pulled towards the prior mean rate r_0/n_0 when the sample size is small. At the other extreme, it will converge to the maximum likelihood estimate when the sample size is large. Thus, the point estimate will reflect uncertainty about the rates even though it does not explicitly represent uncertainty.

It is also possible to obtain an explicit representation of uncertainty using only the representations that we have already posited. The posterior variance is given by:

$$\frac{R_i(t) + r_0}{[N_i(t) + n_0]^2} = \frac{\hat{\lambda}_i(t)}{N'_i(t)}. \quad (27)$$

However, we will not make use of this quantity in what follows.

In this section, we have derived a simple, approximately Bayesian estimation procedure. More sophisticated Bayesian algorithms, such as particle filtering (Daw & Courville, 2008; Gershman et al., 2010) and assumed density filtering (Daw et al., 2008), have been proposed to explain Pavlovian conditioning, albeit in discrete time and with different probabilistic

assumptions. An interesting avenue for future work will be to understand how these more sophisticated algorithms could be put to work for RET.

Deriving Timescale Invariance

Given the learning algorithm described above, we can now examine the relationship between reinforcements to acquisition (R^*) and the informativeness of a Pavlovian protocol (C/T). Plugging Eq. 25 into Eq. 12 yields:

$$1 + \frac{\omega R_{CS}(t)/N_{CS}(t) + (1 - \omega)r_0/n_0}{\omega R_B(t)/t + (1 - \omega)r_0/n_0} > \beta. \quad (28)$$

In the standard delay conditioning protocol, $R_{CS}(t)/N_{CS}(t) \approx 1/T$ and $R_B(t)/t \approx 0$. The approximation will typically be accurate for large t . Putting this together with $\omega = 1 - n_0/(\eta t + n_0)$ and $t = RC$, we can solve for the reinforcements to acquisition:

$$R^* \approx \frac{r_0(\beta - 2)}{\eta} \left(\frac{C}{T} \right)^{-1}. \quad (29)$$

We thus obtain the law of timescale invariance, where number of reinforcements to acquisition is inversely proportional to informativeness on the log scale. Because the term multiplying informativeness does not depend on the experimental protocol, we can treat it as a constant, $k = r_0(\beta - 2)/\eta$, which we will fit into experimental data in the next section.

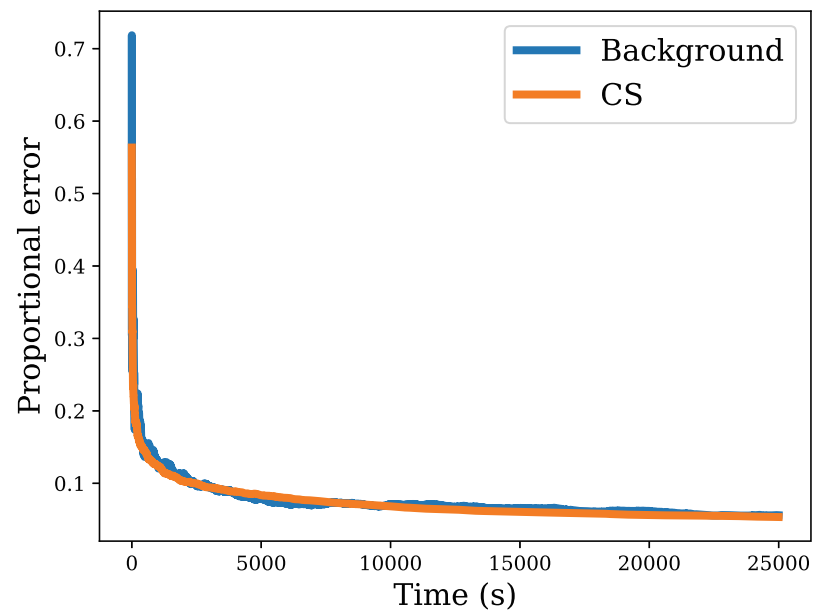
Note that the derived law does not depend on n_0 at all, while r_0 only enters as a scale factor. This means that the choice of the prior parameter values does not fundamentally change this aspect of the model predictions. We will not make strong claims about how to set these values.

Model Validation: Simulations and Data Analysis

Because the theory relies on several approximations, it is important that we validate it using simulations and data analysis. For all simulations, we used the following parameters: $r_0 = 0.1, n_0 = 1, \eta = 0.7$.

First, we show that the approximate learning algorithm correctly estimates rates when they are generated from a stimulus-dependent superposition of Poisson processes, as assumed by the theory. Figure 4 plots the proportional estimation error, $|\hat{\lambda}_i - \lambda_i|/\lambda_i$, over learning for the case where the background rate is 0.5, the CS rate is 1.5, the interstimulus interval is 2, and the intertrial interval is 5. The algorithm accurately learns both rates in a realistic number of trials; the learning curve begins to asymptote after approximately 75 trials (5000 s).

Fig. 4 Estimation error. Reinforcements are generated according to the sum of CS and background Poisson processes. The proportional estimation error for the rates decreases to near 0 over the course of learning



Next, we verified that the model produces timescale invariance in a standard Pavlovian delay conditioning protocol. The left panel of Fig. 5 shows the simulated log-transformed decision variable, $\log \frac{\hat{\lambda}_{CS} + \hat{\lambda}_B}{\hat{\lambda}_B}$, over the course of training, for 3 different interstimulus intervals and a fixed intertrial interval of 48 s. This produces slower learning for longer interstimulus intervals. Data from a pigeon autoshaping experiment (Gibbon et al., 1977), showing the time of acquisition for each condition, is superimposed on the curves.⁸ The decision variable curve intersects these points at roughly the same value across conditions, consistent with the use of a fixed threshold (β). The right panel of Fig. 5 shows the same thing but now for conditions where the intertrial interval is rescaled separately for each condition to maintain a fixed ratio with the interstimulus interval. In this case, all of the curves superimpose—the trajectory of the decision variable is timescale invariant. The empirical acquisition times still intersect these curves at roughly the same value, indicating that the decision threshold has not changed relative to the fixed intertrial interval case. The critical difference is that now these acquisition times cluster together due to timescale invariance.

Finally, we quantitatively evaluate the law derived in the previous section, $R^* = k/(C/T)$, and compare it to a slightly different law proposed by Harris and Gallistel (2024), $R^* = k/(C/T - 1)$. As we've already pointed out, this second law is equivalent to $R^* = k/(I/T)$. We fit the k param-

eter for both laws to several datasets (Gibbon & Balsam, 1981; Balsam et al., 2024; Harris & Gallistel, 2024) by first log-transforming both sides and then minimizing summed squared error (i.e., maximum likelihood estimation under a Gaussian noise model). We then computed the Bayesian information criterion (BIC) approximation of the marginal likelihood (Bishop, 2006):

$$\log P(\text{data}|M) \approx -\frac{1}{2} \text{BIC}(M), \quad (30)$$

where M indexes models and “data” refers to the set of measured reinforcements to acquisition. We converted this into a posterior probability (assuming a uniform prior over models):

$$P(M_1|\text{data}) = \frac{P(\text{data}|M_1)}{P(\text{data}|M_2) + P(\text{data}|M_1)}. \quad (31)$$

Figure 6 shows the curve fits for each dataset. Although the models make very similar predictions, the Bayesian model comparison decisively favored the C/T law rather than the I/T law (all posterior probabilities were greater than 0.99). This result thus confirms that the predictions of the model are empirically well-grounded.

Conclusion

All theories of learning make claims about both learning (what information is extracted from sensory inputs for storage in memory) and performance (how learned information is mapped to behavior). Classical associative theories, exemplified by the Rescorla-Wagner model, claim that learn-

⁸ The criterion used in the Gibbon et al. analyses (an absolute measure of responding) is problematic as a measure of acquisition speed because it is potentially confounded with asymptotic rates of responding. This issue does not apply to the other datasets, where relative measures of responding were used.

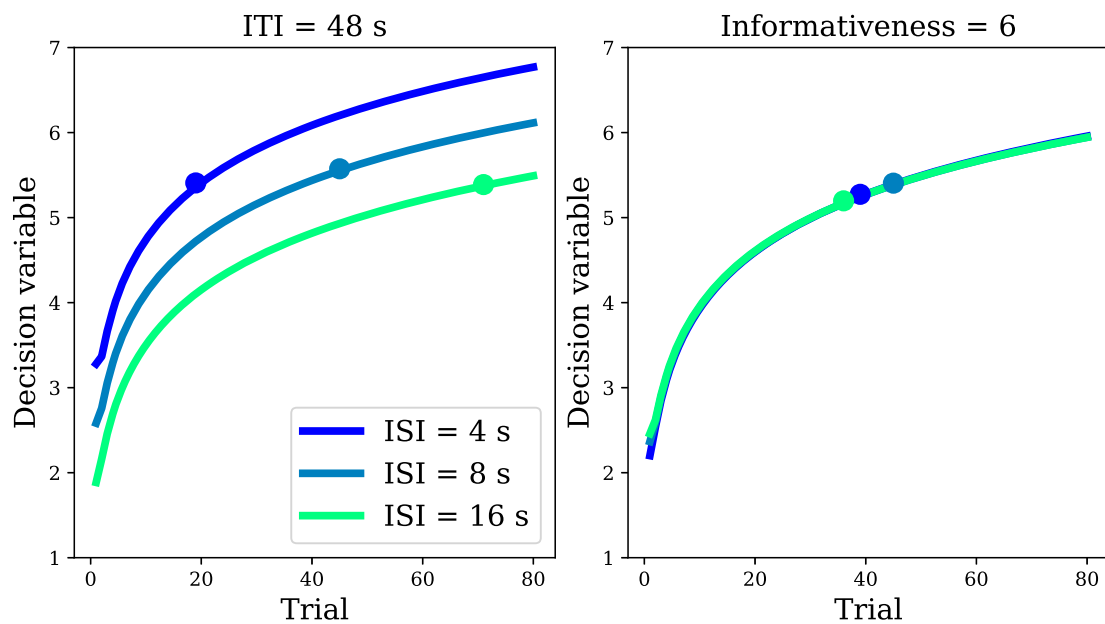


Fig. 5 Timescale invariance of delay conditioning. Each plot shows the simulated log-transformed decision variable, $\log(\hat{\lambda}_{CS} + \hat{\lambda}_B) - \log(\hat{\lambda}_B)$, as a function of trial in a standard Pavlovian delay conditioning protocol. The colored dots show trials (i.e., reinforcements) to acquisition from pigeon autoshaping experiments (Gibbon et al., 1977), as analyzed

in Gallistel and Gibbon (2000). The dots have been superimposed on the curves to facilitate comparison. (Left) The intertrial interval (ITI) is fixed at 48 s, while the interstimulus interval (ISI) varies across conditions. (Right) The ITI is rescaled to maintain a fixed ratio with the ISI

ing is association formation and performance (conditioned responding in a Pavlovian context) is a read-out of associative strength.⁹ Modern representational theories, exemplified by RET, claim that learning is rate estimation and performance is a read-out of contingency between stimuli. It would appear that there is little common ground between these theories, but that appearance is misleading. This paper has shown that a learning algorithm closely resembling the Rescorla-Wagner model can be used to estimate rates—a computationally practical alternative to the algorithmic solution proposed by Gallistel (1990).

A critical difference between the theories concerns performance. As recognized by a number of different theoretical traditions, not all of them representational in the manner of RET (see, for example, Miller & Matzel, 1988; Bouton, 1993; Stout & Miller, 2007), conditioned responding is fundamentally comparative in nature, influenced by CS and US properties that go beyond their direct association. RET formalizes this idea by positing that conditioned responses are generated when the stimulus-conditional rate estimate exceeds the background rate estimate by some threshold. Another contribution of this paper is to place this proposal on firmer theoretical footing, showing how a probabilis-

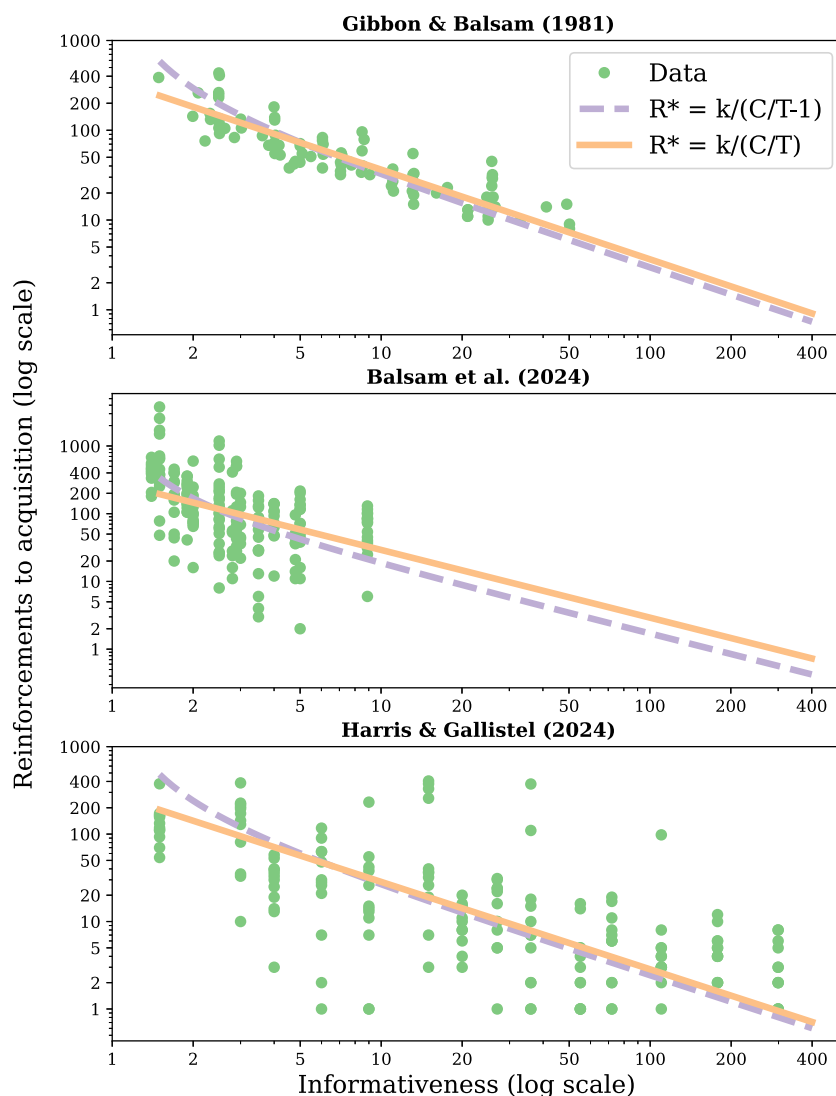
tic approach to rate estimation, in combination with an information-gain decision rule, can (approximately) yield timescale invariance, a fundamental empirical law of Pavlovian delay conditioning.

The model makes a number of predictions that could be tested in future experiments. First, changing the prior should affect the number of trials to acquisition (though not timescale invariance). An animal exposed to multiple different conditioning protocols may come to develop an informative prior by tuning the r_0 and n_0 parameters. For example, if the protocols tend to have a high total number of reinforcements, then r_0 should be higher, which (according to Eq. 29) should slow learning down. Intuitively, the animal should expect higher rates of reinforcement both during and between CS presentations, which will push the decision variable closer to unity. Second, the total amount of prior experience, which should affect n_0 , does not enter into Eq. 29; this implies that it should have no effect on learning speed.

One reason why contiguity-based theories like the Rescorla-Wagner model (and its descendants) have had such staying power is that they fit snugly with modern neurobiological theories of learning based on Hebbian synaptic plasticity. From Gallistel's viewpoint, this is a fundamental error, because both of these theories are irredeemably flawed (Gallistel & Matzel, 2013). While there are many reasons to endorse such a viewpoint (see also Gershman, 2023), we shouldn't throw

⁹ See Honey et al. (2020) for an example of an associative theory that makes more complex assumptions about the mapping from associations to behavior.

Fig. 6 Informativeness predicts learning speed. Each panel shows a measure of learning speed (number of reinforcements prior to the satisfaction of a learning criterion) as a function of informativeness (the ratio of the average US-US interval to the average CS-US interval) on a logarithmic scale. The learning criterion is defined differently across the studies. Gibbon and Balsam (1981) assume learning has occurred when the animal produces a conditioned response on 3/4 consecutive trials. Balsam et al. (2024) and Harris and Gallistel (2024) use a more complex response parsing algorithm combined with a statistical threshold for deciding when the conditional response rate exceeds the unconditional response rate



the baby out with the bathwater. The fact that rate estimation can be reduced to a form of error-driven learning suggests that it could be implemented with the machinery thought to exist in the dopamine system. Considerable evidence supports the hypothesis that dopamine reports reward prediction errors (Montague et al., 1996; Glimcher, 2011; Watabe-Uchida et al., 2017), though this hypothesis has had its own share of controversy (Gershman et al., 2024). The rate estimates might be implemented as synaptic strengths (Austen et al., 2021, 2022), or as cell-intrinsic molecular codes (Gallistel, 2017). Which of these possibilities is most plausible will require additional linking assumptions and experimental tests. The important takeaway is that we can begin to build bridges between representational and algorithmic theories of learning—without invoking the concept of association.

Acknowledgements Randy Gallistel provided a helpful discussion of the ideas in this paper. Gautam Rao, Lukasz Kusmierz, and Sandy Doan provided feedback on an earlier draft.

Author Contributions S.G. did everything.

Funding The work was supported by the Air Force Office of Scientific Research grant FA9550-20-1-0413 and a Schmidt Sciences Polymath Award.

Data Availability All data and analysis code is provided in a GitHub repository: https://github.com/sjgershm/rate_estimation.

Code Availability Code for reproducing these analyses is available at https://github.com/sjgershm/rate_estimation.

Declarations

Ethics Approval Not applicable

Consent to Participate Not applicable

Consent for Publication Not applicable

Competing Interests The author declares no competing interests.

References

- Austen, J., Sprengel, R., & Sanderson, D. (2022). Reinforcement rate and the balance between excitatory and inhibitory learning: Insights from deletion of the GluA1 AMPA receptor subunit. *Journal of Experimental psychology. Animal Learning and Cognition*, 48, 307–314.
- Austen, J. M., Pickering, C., Sprengel, R., & Sanderson, D. J. (2021). Dissociating representations of time and number in reinforcement-rate learning by deletion of the GluA1 AMPA receptor subunit in mice. *Psychological Science*, 32, 204–217.
- Austen, J. M., & Sanderson, D. J. (2020). Cue duration determines response rate but not rate of acquisition of Pavlovian conditioning in mice. *Quarterly Journal of Experimental Psychology*, 73, 2026–2035.
- Balsam, P., & Schwartz, A. (1981). Rapid contextual conditioning in autoshaping. *Journal of Experimental psychology. Animal Behavior Processes*, 7, 382–393.
- Balsam, P. D., Fairhurst, S., & Gallistel, C. R. (2006). Pavlovian contingencies and temporal information. *Journal of Experimental psychology. Animal Behavior Processes*, 32, 284–294.
- Balsam, P. D., & Gallistel, C. R. (2009). Temporal maps and informativeness in associative learning. *Trends in Neurosciences*, 32, 73–78.
- Balsam, P. D., Simpson, E. H., Taylor, K., Kalmbach, A., & Gallistel, C. R. (2024). Learning depends on the information conveyed by temporal relationships between events and is reflected in the dopamine response to cues. *Science Advances*, 10, eadi7137.
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- Bouton, M. (1993). Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. *Psychological Bulletin*, 114, 80–99.
- Bouton, M., & Sunsay, C. (2003). Importance of trials versus accumulating time across trials in partially reinforced appetitive conditioning. *Journal of Experimental psychology. Animal Behavior Processes*, 29, 62–77.
- Chapman, G. B., & Robbins, S. J. (1990). Cue interaction in human contingency judgment. *Memory & Cognition*, 18, 537–545.
- Danks, D. (2003). Equilibria of the Rescorla-Wagner model. *Journal of Mathematical Psychology*, 47, 109–121.
- Daw, N., & Courville, A. (2008). The pigeon as particle filter. *Advances in Neural Information Processing Systems*, 20, 369–376.
- Daw, N. D., Courville, A. C., & Dayan, P. (2008). Semi-rational models of conditioning: The case of trial order. In N. Chater & M. Oaksford (Eds.), *The Probabilistic Mind* (pp. 431–452). Oxford University Press.
- Escobar, M., & Miller, R. R. (2004). A review of the empirical laws of basic learning in Pavlovian conditioning. *International Journal of Comparative Psychology*, 17, 279–303.
- Gallistel, C. (2021). Robert Rescorla: Time, information and contingency. *Revista de Historia de la Psicología*, 41, 7–21.
- Gallistel, C., Craig, A., & Shahan, T. (2019). Contingency, contiguity, and causality in conditioning: Applying information theory and Weber's law to the assignment of credit problem. *Psychological Review*, 126(5), 761–773.
- Gallistel, C., & Gibbon, J. (2000). Time, rate, and conditioning. *Psychological Review*, 107(2), 289–344.
- Gallistel, C. R. (1990). *The organization of learning*. MIT Press.
- Gallistel, C. R. (2017). The coding question. *Trends in Cognitive Sciences*, 21, 498–508.
- Gallistel, C. R., Craig, A. R., & Shahan, T. A. (2014). Temporal contingency. *Behavioural Processes*, 101, 89–96.
- Gallistel, C. R., & Latham, P. E. (2023). Bringing Bayes and Shannon to the study of behavioural and neurobiological timing and associative learning. *Timing & Time Perception*, 11, 29–89.
- Gallistel, C. R., & Matzel, L. D. (2013). The neuroscience of learning: Beyond the Hebbian synapse. *Annual Review of Psychology*, 64, 169–200.
- Gamzu, E., & Williams, D. R. (1971). Classical conditioning of a complex skeletal response. *Science*, 171, 923–925.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. CRC Press.
- Gershman, S. J. (2023). The molecular memory code and synaptic plasticity: A synthesis. *Biosystems*, 224, 104825.
- Gershman, S. J., Assad, J. A., Datta, S. R., Linderman, S. W., Sabatini, B. L., Uchida, N., & Wilbrecht, L. (2024). Explaining dopamine through prediction errors and beyond. *Nature Neuroscience*, pp. 1–11.
- Gershman, S. J., Blei, D. M., & Niv, Y. (2010). Context, learning, and extinction. *Psychological Review*, 117, 197–209.
- Gerstenberg, T. (2024). Counterfactual simulation in causal cognition. *Trends in Cognitive Sciences*.
- Ghirlanda, S. (2022). A response function that maps associative strengths to probabilities. *Journal of Experimental psychology. Animal Learning and Cognition*, 48, 161–168.
- Ghirlanda, S., & Ibadullayev, I. (2015). Solution of the comparator theory of associative learning. *Psychological Review*, 122, 242–259.
- Gibbon, J., Baldock, M., Locurto, C., Gold, L., & Terrace, H. (1977). Trial and intertrial durations in autoshaping. *Journal of Experimental Psychology: Animal Behavior Processes*, 3, 264–284.
- Gibbon, J., & Balsam, P. (1981). Spreading association in time. In C. M. In, H. S. Locurto, & J. G. Terrace (Eds.), *Autoshaping and conditioning theory* (pp. 219–253). Academic Press.
- Gibbon, J., Berryman, R., & Thompson, R. L. (1974). Contingency spaces and measures in classical and instrumental conditioning. *Journal of the Experimental Analysis of Behavior*, 21, 585–605.
- Glimcher, P. W. (2011). Understanding dopamine and reinforcement learning: The dopamine reward prediction error hypothesis. *Proceedings of the National Academy of Sciences*, 108(supplement-3), 15647–15654.
- Harris, J., Kwok, D., & Gottlieb, D. (2019). The partial reinforcement extinction effect depends on learning about nonreinforced trials rather than reinforcement rate. *Journal of Experimental psychology. Animal Learning and Cognition*, 45, 485–501.
- Harris, J. A., & Gallistel, C. (2024). Information, certainty, and learning. *eLife*, 13.
- Holland, P. C. (2000). Trial and intertrial durations in appetitive conditioning in rats. *Animal Learning & Behavior*, 28, 121–135.
- Holmes, N., Chan, Y., & Westbrook, R. (2019). A combination of common and individual error terms is not needed to explain associative changes when cues with different training histories are conditioned in compound: A review of Rescorla's compound test procedure. *Journal of Experimental psychology. Animal Learning and Cognition*, 45, 242–256.
- Honey, R., Dwyer, D., & Iliescu, A. (2020). HeiDI: A model for Pavlovian learning and performance with reciprocal associations. *Psychological Review*, 127, 829–852.
- Jenkins, H. (1965). Judgment of contingency between responses and outcomes. *Psychological Monographs*, 79, 1–17.
- Kalmbach, A., Chun, E., Taylor, K., Gallistel, C., & Balsam, P. (2019). Time-scale-invariant information-theoretic contingencies in discrimination learning. *Journal of Experimental psychology. Animal Learning and Cognition*, 45, 280–289.
- Lattal, K. (1999). Trial and intertrial durations in Pavlovian conditioning: Issues of learning and performance. *Journal of Experimental psychology. Animal Behavior Processes*, 25, 433–450.

- Lattal, K., & Abel, T. (2001). An immediate-shock freezing deficit with discrete cues: A possible role for unconditioned stimulus processing mechanisms. *Journal of Experimental psychology: Animal Behavior Processes*, 27, 394–406.
- Lewis, D. (1973). Causation. *Journal of Philosophy*, 70, 556–567.
- Lindblom, L., & Jenkins, H. (1981). Responses eliminated by non-contingent or negatively contingent reinforcement recover in extinction. *Journal of Experimental psychology: Animal Behavior Processes*, 7, 175–190.
- Lindley, D. V. (1956). On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, 27, 986–1005.
- Miguez, G., Witnauer, J., & Miller, R. (2011). The role of contextual associations in producing the partial reinforcement acquisition deficit. *Journal of Experimental psychology: Animal Behavior Processes*, 38, 40–51.
- Miller, R. R., Barnet, R. C., & Grahame, N. J. (1995). Assessment of the Rescorla-Wagner model. *Psychological Bulletin*, 1(17), 3–363.
- Miller, R. R., & Escobar, M. (2001). Contrasting acquisition-focused and performance-focused models of acquired behavior. *Current Directions in Psychological Science*, 10, 141–145.
- Miller, R. R., & Matzel, L. D. (1988). The comparator hypothesis: A response rule for the expression of associations. In *Psychology of Learning and Motivation* (Vol. 22, pp. 51–92). Elsevier.
- Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience*, 16, 1936–1947.
- Overmier, J., Payne, R., Brackbill, R., Linder, B., & Lawry, J. (1979). On the mechanism of the post-asymptotic CR decrement phenomenon. *Acta Neurobiologiae Experimentalis*, 39, 603–620.
- Pearl, J. (2000). *Causality: models, reasoning, and inference*. Cambridge University Press.
- Randich, A., & LoLordo, V. M. (1979). Preconditioning exposure to the unconditioned stimulus affects the acquisition of a conditioned emotional response. *Learning and Motivation*, 10, 245–277.
- Reiss, S., & Wagner, A. R. (1972). CS habituation produces a “latent inhibition effect” but no active “conditioned inhibition”. *Learning and Motivation*, 3, 237–245.
- Rescorla, R. (1967). Pavlovian conditioning and its proper control procedures. *Psychological Review*, 74, 71–80.
- Rescorla, R. (1968). Probability of shock in the presence and absence of CS in fear conditioning. *Journal of Comparative and Physiological Psychology*, 66, 1–5.
- Rescorla, R. (1973). Effect of US habituation following conditioning. *Journal of Comparative and Physiological Psychology*, 82, 137–143.
- Rescorla, R. A. (1988). Behavioral studies of Pavlovian conditioning. *Annual Review of Neuroscience*, 11, 329–352.
- Rescorla, R. A. (2000). Associative changes with a random CS-US relationship. *The Quarterly Journal of Experimental Psychology: Section B*, 53, 325–340.
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. Black & W. Prokasy (Eds.), *Classical Conditioning II: Current Research and theory* (pp. 64–99). New York, NY: Appleton-Century-Crofts.
- Stout, S., & Miller, R. (2007). Sometimes-competing retrieval (SOCR): A formalization of the comparator hypothesis. *Psychological Review*, 114, 759–783.
- Sunsay, C., & Bouton, M. E. (2008). Analysis of a trial-spacing effect with relatively long intertrial intervals. *Learning & Behavior*, 36, 104–115.
- Terrace, H. S., Gibbon, J., Farrell, L., & Baldock, M. (1975). Temporal factors influencing the acquisition and maintenance of an autoshaped keypeck. *Animal Learning & Behavior*, 3, 53–62.
- Thein, T., Westbrook, R., & Harris, J. (2008). How the associative strengths of stimuli combine in compound: Summation and overshadowing. *Journal of Experimental psychology: Animal Behavior Processes*, 34, 155–166.
- Ward, R., Gallistel, C., Jensen, G., Richards, V., Fairhurst, S., & Balsam, P. (2012). Conditioned stimulus informativeness governs conditioned stimulus-unconditioned stimulus associability. *Journal of Experimental Psychology: Animal Behavior Processes*, 38, 217–232.
- Watabe-Uchida, M., Eshel, N., & Uchida, N. (2017). Neural circuitry of reward prediction error. *Annual Review of Neuroscience*, 40, 373–394.
- Yin, H., Grahame, N. J., & Miller, R. R. (1993). Extinction of comparator stimuli during and after acquisition: Differential facilitative effects on Pavlovian responding. *Learning and Motivation*, 24, 219–241.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.