Rate estimation revisited

Samuel J. Gershman
Department of Psychology and Center for Brain Science, Harvard University

July 28, 2024

Abstract

Two views of Pavlovian conditioning have dominated 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 and Gibbon, 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. This suggests that the gulf between associative and representational theories is smaller than previously thought.

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 and 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 and Gallistel, 2009; Ward et al., 2012; Gallistel et al., 2014, 2019; Gallistel and Latham, 2022; Kalmbach et al., 2019). These models have successfully explained a wealth of quantitative data on conditioning, and have also overcome some 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 an information-theoretic decision rule, give rise to an important quantitative empirical law: the timescale invariance of learning, as explained in the next section.

From contiguity to contingency

In most associative theories of learning, temporal contiguity between stimuli is a necessary (and sometimes sufficient) condition for the formation 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, 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 and 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

---

[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.

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 and 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). At time n, the US prediction $\hat{r}(n)$ is a linear combination of CS associative strengths $w_i(n)$ and the CS presence $x_i(n) \in \{0, 1\}$ for CS i:

$$\hat{r}(n) = w_b(n) + \sum_i w_i(n)x_i(n), \tag{1}$$

where b indicates 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. When the US $r(n) \in \{0, 1\}$ is delivered, the associative strength is updated based on the error signal $\delta(n) = r(n) - \hat{r}(n)$:

$$w_i(n + 1) = w_i(n) + \alpha x_i(n)\delta(n), \tag{2}$$

where $\alpha \in [0, 1]$ is a learning rate. 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, $\hat{r}(\infty)$, then acquisition speed will be 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 can be found analytically (Chapman and Robbins, 1990; Gallistel, 1990; Danks, 2003):

$$w_b(\infty) = \frac{\bar{R} - \bar{R}_i}{C - T} \tag{3}$$

$$w_i(\infty) = \frac{\bar{R}_i}{T} - w_b(\infty), \tag{4}$$

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*), $\bar{R}_i$ 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, $\bar{R}_i = \bar{R} = 1$, so that $w_i(\infty) = 1/T$ 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, $\hat{r}(\infty) = w_i(\infty) + w_b(\infty) = 1/T$, which means that the asymptotic US prediction is *not* timescale invariant.

More generally, $\hat{r}(\infty) = \bar{R}_i/T$ 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 and Miller, 2004, for a review). For example, adding unsignaled US deliveries during the intertrial interval (Rescorla, 1968), prior to conditioning (Randich and LoLordo, 1979; Balsam and Schwartz, 1981; Overmier et al., 1979), or after (Rescorla, 1973; Overmier et al., 1979) conditioning, 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 and Bouton, 2008), pre-exposing an animal to the experimental context (Lattal and Abel, 2001), or extinguishing the context following conditioning (Yin et al., 1993; Miguez et al., 2011).

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*. On the other hand, it is unclear whether temporal contiguity is inadequate as a principle of Pavlovian *learning*. 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 and Escobar, 2001). Rate Estimation Theory (RET; reviewed in the next section) falls into this class, as does the sometimes-competing retrieval model (Stout and Miller, 2007) and Rescorla's semi-formal contingency model (Rescorla, 1967). The stimulus relationships are variously interpreted by these models as rates, associations, or probabilities; we will see that mathematically these concepts are closely related. Indeed, 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 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 the relative response rule.

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),[2] then the probability of necessity and sufficiency is given by:

$$\Delta P = P(\hat{r}(\infty) = 1 | x_i = 1) - P(\hat{r}(\infty) = 1 | x_i = 0), \tag{5}$$

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 precisely $\Delta P = w_i(\infty) - w_b(\infty)$, a fact also noticed by Chapman and Robbins (1990). This further implies that ΔP is 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 (see also 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:

$$\lim_{\Delta t \to 0} \frac{\bar{R}_i}{T/\Delta t} - \frac{\bar{R} - \bar{R}_i}{(C - T)/\Delta t} = 0, \tag{6}$$

where division by Δt translate 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. Shoehorning it into discrete-time probability distributions 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 and Gibbon, 2000) is the replacement of discrete timesteps (n) with continuous time (t). Some events, such as US deliveries

---

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

(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 reduce 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, $r(t) \in \{0, 1\}$, where t indexes time. 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:

  $$r(t) = \sum_i r_i(t), \tag{7}$$

  where i indexes stimuli.

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

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

  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). 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

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

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

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:

$$\frac{\tilde{R}_i}{N_i} = \sum_j \lambda_j \frac{N_{ij}}{N_i}, \tag{10}$$

where $\tilde{R}_i = \int_t x_i(t) r(t) dt$ is the total number of reinforcements observed in the presence of stimulus i, $N_i = \int_t x_i(t) dt$ is the cumulative record of stimulus i, and $N_{ij} = \int_t x_i(t) x_j(t) dt$ is the cumulative pairwise record of stimulus i and j (i.e., the total amount of time during which they were presented together). Note that we have dropped the time index from these variables but in practice we assume that the animal observes the process for a finite time (in which case the equation is only approximate).

Gallistel refers to the quantity $u_i = \tilde{R}_i/N_i$ as the "uncorrected" rate estimate. Its expectation equals the true rate λ_i only when stimulus i is consistently presented alone (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:

$$\hat{\lambda} = \mathbf{A}^{-1}\mathbf{u}, \tag{11}$$

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

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 CS i whenever

$$\frac{\hat{\lambda}_i + \hat{\lambda}_b}{\hat{\lambda}_b} > \beta, \tag{12}$$

where β is a threshold parameter and λ_b is the reinforcement rate for the background stimulus (i.e., a constantly present stimulus process, $x_b(t) = 1$ for all t). Later work, starting with Balsam et al. (2006), adopted an information-theoretic version of this decision rule, where the animal responds whenever the estimated *informativeness* of the CS exceeds a threshold (see also Balsam and Gallistel, 2009). Informativeness is defined (up to a constant) as the mutual information between the CS and reinforcement processes, assuming both are Poisson-distributed:

$$H_i = \log \frac{\lambda_i}{\lambda_b} = \mathcal{H}[r_i] - \mathcal{H}[r_i|x_i] + \text{const.} \tag{13}$$

where $\mathcal{H}[r_i]$ is the unconditional entropy of r_i, and $\mathcal{H}[r_i|x_i]$ is the conditional entropy given the stimulus process. Intuitively, informativeness measures how much the animal's uncertainty about upcoming reinforcement is reduced by observing the CS. Because the true rates are unobserved, the animal is assumed to use its rate estimates.[3] Recently, Gallistel and Latham (2022) have formulated a related decision criterion that takes into account the animal's uncertainty about the rates (more on this later).

This family of relative response rules 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 assume 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. These phenomena, which increase $\hat{\lambda}_b$, are naturally explained by the relative response rule. The relative response rule is also at the heart of how RET explains timescale invariance.

Explaining timescale invariance

Gallistel and Gibbon (2000) derived timescale invariance with the RET framework as follows. Recall that in this setting we are considering a single CS i and the background stimulus b, where the US always follows the CS after a fixed delay. Because of rate additivity, we have $\lambda_i + \lambda_b = T^{-1}$, where T is the CS-US interval. Gallistel and Gibbon further assumed that $\lambda_b = (RI)^{-1}$, where I is the intertrial interval and R is the number of trials (equivalent 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 stimates ($\hat{\lambda} = \lambda$), RET predicts that a conditioned response will be produced when

$$\frac{RI}{T} > \beta, \tag{14}$$

or equivalently when

$$R > \beta \frac{T}{I}. \tag{15}$$

We thus recover the important empirical law that trials (reinforcements) to acquisition (R) is linearly related to the ratio of interstimulus and intertrial intervals, such that rescaling both intervals by a constant leaves the acquisition speed unchanged (Gibbon et al., 1977).

While this 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.[4]

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

---

[3] Note that informativeness (using the rate estimates) is linearly related to the original ratio criterion.

[4] Gallistel and Latham (2022) develop 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.

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 shortcoming 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 how an information-theoretic 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 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:

$$\frac{d\hat{\lambda}_i(t)}{dt} = \frac{x_i(t)[r_i(t) - \hat{\lambda}_i(t)]}{N_i(t)}, \tag{16}$$

where $N_i(t) = \int_0^t x_i(\tau)d\tau$ is the cumulative record of stimulus i.

Under this update procedure, $\hat{\lambda}_i(t) = R_i(t)/N_i(t)$, where $R_i(t) = \int_0^t x_i(\tau)r_i(\tau)d\tau$ is the cumulative record of reinforcements in the presence of stimulus i. If we only track $\hat{\lambda}_i(t)$ and $N_i(t)$, we can always recover the cumulative record of reinforcements by the transformation $R_i(t) = N_i(t)\hat{\lambda}_i(t)$.

Estimation when the components are unobserved

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

$$r_i(t) \approx \delta(t) + \hat{\lambda}_i(t)x_i(t), \tag{17}$$

where

$$\delta(t) = r(t) - \sum_j \hat{\lambda}_j(t)x_j(t) \tag{18}$$

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

$$\mathbb{E}[r_i(t)] = \mathbb{E}[\delta(t)] + \lambda_i x_i(t), \tag{19}$$

where $\mathbb{E}[\delta(t)] = 0$. Plugging the approximation into Eq. 16 yields:

$$\begin{aligned}\frac{d\hat{\lambda}_i(t)}{dt} &\approx \frac{x_i(t)[\delta(t) + \hat{\lambda}_i(t)x_i(t) - \hat{\lambda}_i(t)]}{N_i(t)} \\ &= \frac{x_i(t)\delta(t)}{N_i(t)}.\end{aligned} \tag{20}$$

This expression is remarkably like 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. Thus, the apparently sharp dichotomy between rate estimation and error-driven learning, as emphasized by Gallistel and Gibbon (2000), is less sharp than previously believed. The critical wedge between these approaches to conditioned behavior is really the decision process rather than the learning process, as discussed later.

Note that this algorithm 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.

Bayesian estimation

The update rules 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 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 cale parameter can be interpreted as the effective elapsed time. 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:

$$\frac{d\hat{\lambda}_i(t)}{dt} = \frac{x_i(t)\delta(t)}{N'_i(t)}, \tag{21}$$

where $N'_i(t) = N_i(t) + n_0$. We will refer to this learning rule as *online Bayesian RET*.

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 $r_0 \to 0, n_0 \to 0$. 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:

$$\hat{\lambda}_i(t) = \frac{R_i(t) + r_0}{N'_i(t)} = \omega\frac{R_i(t)}{N_i(t)} + (1-\omega)\frac{r_0}{n_0}, \tag{22}$$

where

$$\omega = \frac{N_i(t)}{N'_i(t)} \tag{23}$$

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 rate 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:

$$\frac{R_i(t) + r_0}{[N_i(t) + n_0]^2} = \frac{\hat{\lambda}_i(t)}{N'_i(t)}. \tag{24}$$

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

Information-theoretic decision rule

To map rate estimates to conditioned responses, we develop an information-theoretic decision rule similar to the one proposed by Balsam et al. (2006). Importantly, this new rule takes into account rate uncertainty.

Given the Gamma posterior over the rates, we can analytically obtain the posterior mean $\hat{H}_i(t)$ and variance $V_i(t)$ of the informativeness:

$$\hat{H}_i(t) = \psi(N_i(t)\hat{\lambda}_i(t)) - \psi(t\hat{\lambda}_b(t)) - \log N_i(t) + \log t \tag{25}$$

$$V_i(t) = \psi_1(N_i(t)\hat{\lambda}_i(t)) + \psi_1(t\hat{\lambda}_b(t)), \tag{26}$$

where $\psi(\cdot)$ is the digamma function, and $\psi_1(\cdot)$ is the trigamma function. Note that we have assumed here statistical independence between the CS and background rates.

One way to define a decision rule is to initiate responding when the animal is confident that informativeness has exceeded a criterion value, β. By moment-matching the posterior over H_i to a Normal distribution, we can approximate the posterior probability that $H_i > \beta$, what we will refer to as the *exceedance probability*, denoted by $\mu(t)$:

$$\mu(t) \approx \Phi\left(\frac{\hat{H}_i(t) - \beta}{\sqrt{V_i(t)}}\right), \tag{27}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard Normal distribution. We hypothesize that the conditiond response rate tracks the exceedance probability. While this changes smoothly with training (see Harris, 2022), it can also be highly non-linear, reflecting the relatively sharp increases in responding characteristically elicited by conditioning protocols (Gallistel et al., 2004). In particular, the rate of change in the learning curve is maximal at $\hat{H}_i(t) = \beta$.

With this setup, we can analyze the trials to acquisition, a measure of (inverse) stimulus associability (Balsam and Gallistel, 2009). We will define trials to acquisition as the number of reinforcements $R(t)$ at which $\hat{H}_i(t) = \beta$, the major inflection point in the learning curve.

Using the approximation $\psi(x) \approx \log x$,[5] we have:

$$\hat{H}_i(t) \approx \log \frac{\hat{\lambda}_i(t)}{\hat{\lambda}_b(t)} = \log \frac{\hat{R}_i(t)t}{\hat{R}_b(t)N_i(t)}. \tag{28}$$

---

[5] The digamma function behaves asymptotically as $\psi(x) \approx \log x - \frac{1}{2x}$. For $x > 3$, the second term contributes negligibly, and the logarithmic approximation is reasonably accurate.

Let $C = t/R(t)$ denote the average cycle time between reinforcements and $T = N_i(t)/R(t)$ denote the average CS-US delay; we then have the relationship $C/T = t/N_i$. Plugging this into Eq. 28:

$$\hat{H}_i(t) \approx \log \frac{\hat{R}_i(t)}{\hat{R}_b(t)} + \log \frac{C}{T}. \tag{29}$$

We now fix t at the time t^* when $\hat{H}_i(t^*) = \beta$ and drop the time index. Solving for reinforcements to acquisitions, R, yields:

$$R \approx e^{\beta} \hat{R}_b \left(\frac{T}{C} + 1 \right) = e^{\beta} \hat{R}_b \frac{T}{I}. \tag{30}$$

This expression approximately recapitulates the timescale invariance of conditioning, $R \propto T/I$, provided $\hat{R}_b$ is constant as a function of T/I. Since $\hat{R}_b = \hat{\lambda}_b(t^*)t^*$, it suffices that $\hat{\lambda}_b(t^*) \propto 1/t^*$. This condition is satisfied if early during training (when there is still appreciable uncertainty about the CS and background rates) a fixed number of reinforcements are attributed to the background, after which the CS rate is high enough to take credit for all subsequent reinforcements. Note, however, that this is an essentially *ad hoc* assumption similar to the one made by Gallistel and Gibbon (2000). In the next section, we show that in practice this model produces near-perfect timescale invariance.

Simulations

Because the theory relies on several approximations, it is important that we validate it using simulations. Code for reproducing these simulations is available at `https://github.com/sjgershm/rate_estimation`.

First, we show that the approximate learning scheme correctly estimates rates when they are generated from a stimulus-dependent superposition of Poisson processes, as assumed by the theory. Figure 1 plots the estimation error over learning for the case where background rate is 0.5, the CS rate is 2, the interstimulus interval is 2, and the intertrial interval is 5. The online Bayesian RET algorithm (with $r_0 = 1$ and $n_0 = 1$) accurately learns both rates in a realistic number of trials; the learning curve begins to asymptote after approximately 75 trials.

Next, we verified that the model produces timescale invariance in a standard Pavlovian delay conditioning protocol. Because conditioned responding is determined by comparison of estimated informativeness with a fixed threshold, it suffices to show that estimated informativeness is constant across variations in interstimulus and intertrial intervals, as long as their ratio is held fixed. As shown in Figure 2, the posterior mean informativeness is indeed almost constant as a function of the interstimulus interval when the ratio is held fixed. Different points on the curve correspond to different scalar multiples of the ratio. The two curves show two different ratios, with greater estimated informativeness when the ratio is larger (i.e., when the intertrial interval is large relative to the interstimulus interval).

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

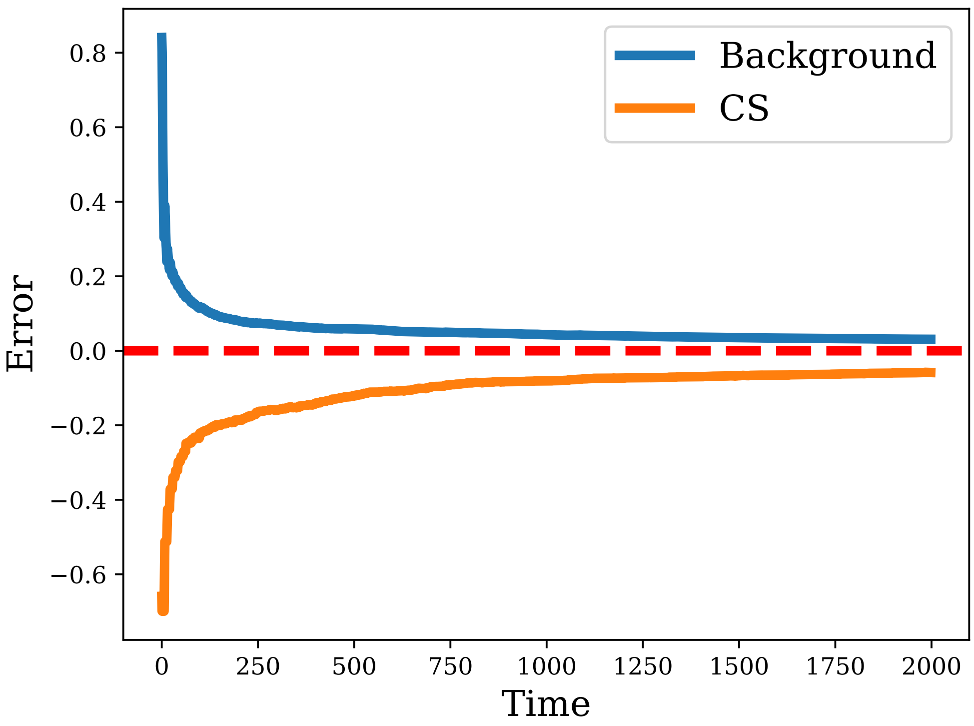

Figure 1: **Learning curve**. Reinforcements are generated according to the sum of CS and background Poisson processes. The estimation error for the rates decreases to near 0 over the course of learning.

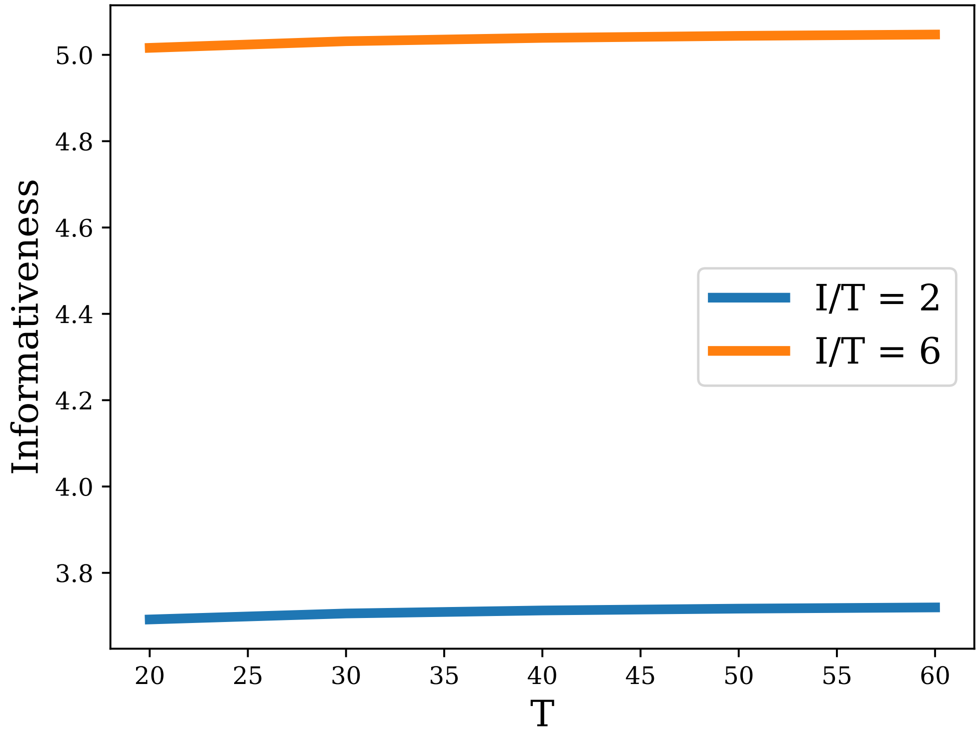

Figure 2: **Timescale invariance of conditioning**. The y-axis shows the posterior mean estimate of informativeness after 60 trials. Informativeness is plotted as a function of the interstimulus interval (T). The two curves correspond to different ratios of intertrial and interstimulus intervals.

behavior). Classical associative theories, exemplified by the Rescorla-Wagner model, claim that learning is association formation and performance (conditioned responding in a Pavlovian context) is a read-out of associative strength.[6] 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).

The 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 and Matzel, 1988; Bouton, 1993; Stout and 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 fully probabilistic approach to rate estimation, in combination with an information-theoretic decision rule, can (approximately) yield timescale invariance, a fundamental empirical law of Pavlovian delay conditioning.

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 and Matzel, 2013). While there are many reasons to endorse such a viewpoint (see also Gershman, 2023), we shouldn't throw 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 important takeway is that we can begin to build bridges between representational and algorithmic theories of learning—without invoking the concept of association.

Acknowledgments

Randy Gallistel provided helpful discussion of the ideas in this paper, and Gautam Rao provided feedback on an earlier draft. The work was supported by the Air Force Office of Scientific Research grant FA9550-20-1-0413.

References

Balsam, P. and Schwartz, A. (1981). Rapid contextual conditioning in autoshaping. *Journal of Experimental psychology. Animal Behavior Processes*, 7:382–393.

Balsam, P. D., Fairhurst, S., and Gallistel, C. R. (2006). Pavlovian contingencies and temporal information. *Journal of Experimental psychology. Animal Behavior Processes*, 32:284–294.

---

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

Balsam, P. D. and Gallistel, C. R. (2009). Temporal maps and informativeness in associative learning. *Trends in Neurosciences*, 32:73–78.

Bouton, M. (1993). Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. *Psychological Bulletin*, 114:80–99.

Chapman, G. B. and 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.

Escobar, M. and 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., and 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. and 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., Craig, A. R., and Shahan, T. A. (2014). Temporal contingency. *Behavioural Processes*, 101:89–96.

Gallistel, C. R., Fairhurst, S., and Balsam, P. (2004). The learning curve: implications of a quantitative analysis. *Proceedings of the National Academy of Sciences*, 101:13124–13131.

Gallistel, C. R. and Latham, P. E. (2022). Bringing Bayes and Shannon to the study of behavioural and neurobiological timing and associative learning. *Timing & Time Perception*, 11:29–89.

Gallistel, C. R. and Matzel, L. D. (2013). The neuroscience of learning: beyond the Hebbian synapse. *Annual Review of Psychology*, 64:169–200.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and 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., and Wilbrecht, L. (2024). Explaining dopamine through prediction errors and beyond. *Nature Neuroscience*, pages 1–11.

Gerstenberg, T. (2024). Counterfactual simulation in causal cognition. *Trends in Cognitive Sciences*.

Gibbon, J., Baldock, M., Locurto, C., Gold, L., and Terrace, H. (1977). Trial and intertrial durations in autoshaping. *Journal of Experimental Psychology: Animal Behavior Processes*, 3:264–284.

Gibbon, J., Berryman, R., and 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. A. (2022). The learning curve, revisited. *Journal of Experimental Psychology: Animal Learning and Cognition*, 48:265–280.

Holland, P. C. (2000). Trial and intertrial durations in appetitive conditioning in rats. *Animal Learning & Behavior*, 28:121–135.

Honey, R., Dwyer, D., and 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., and 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. and 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.

Miguez, G., Witnauer, J., and 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. and Escobar, M. (2001). Contrasting acquisition-focused and performance-focused models of acquired behavior. *Current Directions in Psychological Science*, 10:141–145.

Miller, R. R. and Matzel, L. D. (1988). The comparator hypothesis: A response rule for the expression of associations. In *Psychology of Learning and Motivation*, volume 22, pages 51–92. Elsevier.

Montague, P. R., Dayan, P., and 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., and 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. and LoLordo, V. M. (1979). Preconditioning exposure to the unconditined stimulus affects the acquisition of a conditioned emotional response. *Learning and Motivation*, 10:245–277.

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. and Wagner, A. R. (1972). A theory of of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In Black, A. and Prokasy, W., editors, *Classical Conditioning II: Current Research and theory*, pages 64–99. Appleton-Century-Crofts, New York, NY.

Stout, S. and Miller, R. (2007). Sometimes-competing retrieval (SOCR): a formalization of the comparator hypothesis. *Psychological Review*, 114:759–783.

Sunsay, C. and 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., and Baldock, M. (1975). Temporal factors influencing the acquisition and maintenance of an autoshaped keypeck. *Animal Learning & Behavior*, 3:53–62.

Ward, R., Gallistel, C., Jensen, G., Richards, V., Fairhurst, S., and 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., and Uchida, N. (2017). Neural circuitry of reward prediction error. *Annual Review of Neuroscience*, 40:373–394.

Yin, H., Grahame, N. J., and Miller, R. R. (1993). Extinction of comparator stimuli during and after acquisition: Differential facilitative effects on Pavlovian responding. *Learning and Motivation*, 24:219–241.