Bringing Bayes and Shannon to the Study of Behavioral and Neurobiological Timing and Associative Learning

C.R. Gallistel
Peter Latham

Abstract

Bayesian parameter estimation and Shannon's theory of information provide tools for analyzing and understanding data from behavioral and neurobiological experiments on interval timing—and from experiments on Pavlovian and operant conditioning, because timing plays a fundamental role in associative learning. In this tutorial, we explain basic concepts behind these tools and show how to apply them to estimating, on a trial-by-trial, reward-by-reward and response-by-response basis, important parameters of timing behavior and neurobiological manifestations of timing in the brain. These tools allow us to assess the trade-off between acting as an ideal observer should act and acting as an ideal agent should act, which is also known as the trade-off between exploration (information gathering) and exploitation (information utilization). The Supplementary Material provides well-documented Matlab™ and Python code that does the basic analyses and graphs the results. A GitHub site accepts further developments of relevant code.

Information theory and Bayesian approaches to statistics are natural companions. Together, they can assist us in analyzing, intuitively understanding, and formally modeling results from experiments in which we investigate the role of interval timing in behavior and in associative learning. It as has long been clear that associative learning depends on timing (C. R. Gallistel & Gibbon, 2000; Gibbon & Balsam, 1981; Stout & Miller, 2007; Yin, Barnet, & Miller, 1994). In this paper, we lay out the basics of Bayesian parameter estimation and Shannon's theory of information, as they apply to the behavioral and neurobiological study of timing and associative learning. Then, we show how to turn this mathematics into useful tools.

Parameter estimation—for example, estimating the means and standard deviations of two distributions— is the first step in data processing. Frequentist approaches to parameter estimation require the collection of samples of pre-specified size. Bayesian parameter estimation naturally applies datum-by-datum, that is response-by-response and reinforcement-by-reinforcement. That makes it a powerful tool in estimating learning rates—how soon timing-based changes in behavior and/or in neurobiological activity appear, and how soon evidence of timed responses appears.

The *entropy difference* (denoted $\triangle H$) between two distributions is the *mutual information* when those distributions are accurately known and suitably discretized. We show that it is useful even when the distributions are not accurately known but assumed to have the form dictated by the *maximum entropy principle.* This principle is an information-theoretic realization of Occam's razor (Jaynes, 1957, 2003).

An example of two distributions assumed in analyzing data from Pavlovian timing experiments are the distribution of inter-reward intervals in the presence of *conditional stimuli* (denoted CSs, for example, a noise that comes on and off unpredictably) and the same distribution in the *context* in which the CS occurs (typically, a test chamber). In our analyses, we assume them to be exponential even when we know they are not and cannot be (for example, when we know they are mixture distributions).

An example from reinforcement learning experiments (aka operant conditioning) is the distribution of inter-response intervals and the distribution of inter-reward intervals. We show that $\triangle H$ is a generally applicable measure of the extent to which two events or two states are associated in time. It applies in many circumstances where the conventional measure of association—the correlation coefficient— cannot be computed (C. R. Gallistel, 2021): It can be computed even when $n = 1$; and it does not presume a linear relationship (Kinney & Atwal, 2014). It has most of the properties of mutual information but not those properties that depend on the assumption of the form of the assumed distribution (for example, the property of being invariant under change in variable).

A second fundamental quantity in information theory is the *Kullback-Leibler divergence*, denoted by Dkl. It measures the extent to which a distribution of interest diverges from a reference distribution. It is a measure the strength of the evidence that the two distributions differ, with possible implications for the neurobiology of memory. It gives the mnemonic cost (in bits) of encoding a datum coming from one distribution, for example, the distribution of waits for reinforcement conditioned on a CS, on the assumption that they come from a reference distribution, for example, the unconditional waits for reinforcement when in the test chamber. The *cumulative* cost of coding the n conditional data already seen is, on average, simply nDkl, the number of data seen times the estimated divergence of the conditional distribution from the unconditional distribution. The nDkl is to $\triangle H$ as the significance of a correlation coefficient is to the coefficient itself: $\triangle H$ measures the association, while the nDkl measures the strength of the evidence for it.

Both $\triangle H$ and the nDkl are computed from estimates of the parameters of the distributions from which the data are assumed to come. These distributions are assumed to be exponential, whether they are or not. This strong simplifying assumption is a major reason for distinguishing $\triangle H$ from mutual information. It has three justifications:

- It makes $\triangle H$ and nDkl computable by simple closed-form formulae.
- There is extensive experimental evidence that the learning rate and the difference in performance in associative protocols is primarily determined by the ratio of reinforcement rates (the reciprocals of the mean waits for reinforcement). This implies that the only statistic that matters to the subject in making these decisions is the rate of reinforcement. Put another way, the behaviorally relevant *sufficient statistics* from a sample of temporal intervals are the number of intervals in the sample and the duration over which these intervals have been observed.

- The first few intervals in a sample provide the lion's share of the information required to estimate the mean interval, but give only a weak and unreliable estimate of the variance. Therefore, they provide little basis for deciding even between the exponential and the Normal as a model for the source of the data.
- When only the estimate of the mean is available, the *maximum entropy principle* (Jaynes, 1957, 2003) dictates the assumption of the exponential form for the source distribution. It is the weakest possible assumption.

Bayesian parameter estimation is essential when *n*'s are small, because integration over the posterior distributions on the parameter estimates takes into account the large uncertainties in estimates made from very small samples. Bayesian parameter estimation supplies the required posterior distributions.

The nDkl is a simple, maximally powerful datum-by-datum measure of the strength of the evidence that a parameter of the distribution of a behavioral or neurobiological variable (for example the response rate) has changed. It allows us to address questions such as, How many reinforced CSs are required for a subject to detect the temporal association between a CS and a US or between a response and a reinforcement? The use of this datum-by-datum measure obviates the need to rely on arbitrary decision criteria such as the number of trials successive trials on which a response is observed. These criteria often demonstrably underestimate the subject's sensitivity to differences and changes in rates of responding (the reciprocals of average wait durations), probabilities and contingencies.

Different evidentiary decision variables—for examples, p values, odds ratios, and nDkl's—are monotonically related because a useful measure must depend monotonically on the information provided by the data. We provide a simple formula that maps from nDkl to p.

Bayesian Parameter Estimation

Traditional statistics at the applied level are based on maximum likelihood estimates of population parameters given a sample—and, usually also on the central limit theorem, which states that sample means will be normally distributed more or less regardless of the form of the distribution from which samples are drawn. In their rigorous application, these measures require one to specify sample sizes in advance of collecting the data. This has led to insistence on a pre-registration of one's experimental protocol, in which one specifies the sample sizes in advance and the inferential statistics to be performed.

These traditional approaches do not work well with small samples unless the effect of one's experimental manipulation is big. However, one often does not know the size of the effect one should expect. One commonly hopes to learn from a proposed experiment whether there is an effect and if so, how big. In that case, specifying sample size in advance is antithetical to the purpose of the experiment.

Moreover, we often want to measure the strength of the evidence as the data come in – that is, as the sample size grows – because the bigger the effect, the more rapidly strong evidence for it emerges and the sooner we can stop the experiment. The slope of the nDkl when plotted as a function of *n* is a measure of effect size; the greater the divergence, the steeper the slope.

Finally, because we are interested in acquisition and extinction and, more generally, in the course of behavioral change, we often want stimulus parameter estimates and behavioral parameter estimates when there are very little data. An example we will treat is when the only datum is the amount of time elapsed before the occurrence of the first response and the first reinforcement in an operant conditioning protocol.

From a subject's perspective, the protocoled events it experiences in our experiments are manifestations of a stochastic process whose form and parameters must be inferred from the observable outcomes the process generates. The evidence for the form and the value of its parameter vector grows stronger as more events are experienced, leading eventually to the appearance of an appropriately timed anticipatory response. We want to compute the strength of the evidence for the form (e.g., exponential or Normal) and its parameter values (e.g., means and variances) as a function of time elapsed and the numbers of relevant events. We want then to plot the strength of the evidence for the behavioral change against the strength of the evidence the subject has about the process that generates the subject's experiences. This enables us to answer the question, How much evidence is required before anticipatory behavior appears?

In Bayesian parameter estimation, one puts a prior distribution on the plausible values for the parameter(s) of the distribution that one believes approximately describes (or will describe) the data. We refer to distributions that describe the data as *source* distributions to distinguish them from *prior* distributions. What we call the source distribution is often called the likelihood; our reasons for calling for using this non-standard terminology are explained later.

The distinction between the source distribution and the prior distributions is fundamental—and often confusing to the uninitiated. Before clarifying it, we cover the basics of distribution functions. They are often not stressed in the statistics education many of us received.

Distributions

Distributions are functions that map from the members of a *support set* to the members of a set of *probabilities* or *probability densities*. In a plot of a distribution, the support set is composed of the possible values a datum might assume, arrayed along the *x* axis. When the support is discrete (in technical language, finite or countably infinite), the distribution assigns *probabilities* to those possibilities (Figure 1). When the support is continuous (in technical language, uncountably infinite), the distribution assigns *probability densities* (Figure 2).

To every probability distribution (think histogram), there corresponds a cumulative probability distribution. The *cumulative distribution* is the cumulative sum (or integral) of the probabilities (or probability densities) as one moves from left to right along the support axis, from the smallest possibility to the largest. As can be seen in the second rows of Figures 1 and 2, cumulative distributions asymptote to 1. That's because the total mass of probability in a distribution must be 1, since its support is a (possibly uncountably infinite) set of mutually exclusive and exhaustive possibilities. Note also that for every cumulative distribution there is a probability distribution, which is found by taking a difference (for discrete distributions) or a derivative (for continuous ones).

A continuous distribution assigns *probability densities* to the members of the support set rather than probabilities (Figure 2). Whereas *probabilities* always fall between 0 and 1 (Figure 1 and Figure 2 bottom row), *probability densities* (Figure 2, top row) may take on values from 0 to + infinity. When, for example, the cumulative probability function is a step from 0 to 1 at some point along the x axis (Figure 2, bottom left), the derivative at the step is infinite, and everywhere else it is 0. This derivative is the unit impulse; it is the limit of a rectangle whose width goes to 0 as its height goes to infinity while maintaining an area of 1 (the total mass of probability in any probability distribution).

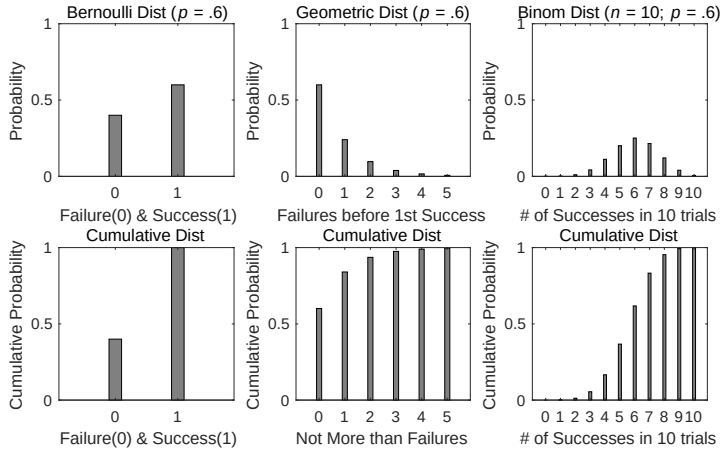

Figure 1. *Three common discrete distributions: the Bernoulli, the geometric and the binomial. They are plotted with bars rather than curves because the support is discrete. Discrete support may always be represented by the integers, as for example, in the common practice of representing "failure" by the integer '0' and "success" by the integer '1' in the support for the Bernoulli distribution. The cumulative probabilities in the bottom row are obtained by moving rightward from bar to bar in the top row, summing the successive probabilities. The geometric distribution may be thought of as the discrete analog of the exponential distribution and the binomial may be thought of as the discrete analog of the Normal, because the exponential and the Normal are the distributions that emerge as the set of possibilities becomes uncountably infinite (as the bars become ever narrower and more numerous.*

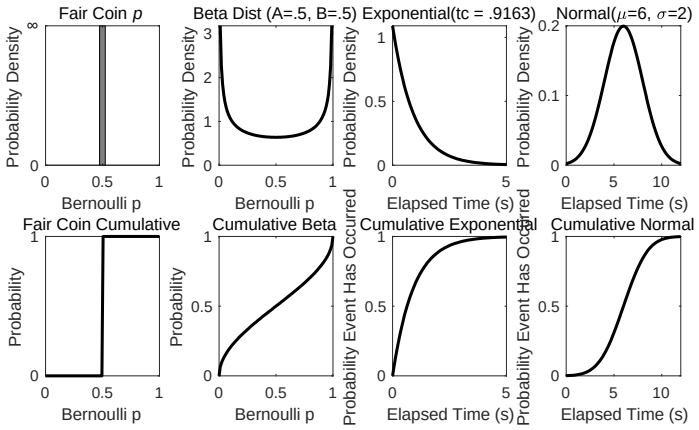

Figure 2. *Four common distributions over continuous variables. It is stipulative that a fair coin have a probability of heads of exactly 0.5; therefore, the distribution of the probability of obtaining heads when flipping a fair coins is a vertical line at p = .5 with no width, infinite height and an area (width x height) = 1. The cumulative distribution for the p values of fair coins is a step from 0 to 1 at 0.5. The beta distribution is a commonly used prior distribution on the Bernoulli p in Bayesian statistics. It has two parameters, which may assume values between 0 and + infinity. The example here uses A = B = .5. These are the values for the so-called Jeffreys prior on the Bernoulli. The probability densities at both extremes become infinite, but, like all proper distributions, the beta distribution integrates to 1 (bottom row). The exponential describes the distribution of the intervals between randomly scheduled events. The support for distributions over continuous variables like interval duration is said to be uncountably infinite because of Cantor's famous proof that there are uncountably many different intervals within any finite interval, no matter how small that finite interval is.*

Distribution functions are determined by their *mathematical form* and by the *values of their parameters*. The form defines a family of distributions. The members of that family are distinguished by the values chosen (or estimated) for their parameters. Thus, for example, a Normal distribution is a family and a Normal distribution for which a mean and standard deviation have been specified is a member of that family.

A distribution with a given form may be parameterized in different ways. This becomes important in Bayesian analysis. For example, the Bernoulli and the geometric distributions may both be parameterized either by p (the probability of a success) or by the *odds* of a success, $p/(1-p)$. Statisticians prefer the former parameterization; bookies prefer the latter. The exponential may be parameterized either by the rate at which events occur, λ, or by the average interval between them, $\mu = 1/\lambda$. The Normal may be parameterized by its mean (μ) and standard deviation (σ), or by its mean and variance (σ^2), by its mean and precision ($\tau = 1/\sigma^2$), or even by its mean and coefficient of variation (σ/μ).

Having covered the basics of probability distributions, we can explain the two distributions that are always in play in Bayesian parameter estimation, the source distribution, which is supported by possible values for the data, and the prior/posterior distribution, which is supported by possible values for the parameter(s) of the source distribution. The source distribution and its prior distribution are distinguished by their support not by their form; however, generally speaking, they also have different forms.

Source distributions

Stochastic models for data may have any number of parameters. In deep learning models, they have millions, even billions. However, the source distributions commonly used in modeling behavior have only 1 or 2 parameters. For example, for Normal distributions, $\theta = [\mu \quad \sigma]$, whereas for the Bernoulli, $\theta = p$; and for the exponential, $\theta = \lambda$, the rate parameter.

Prior Distributions

The support for a *prior* is the parameter vector of the source—*not the possible values for a datum*. Thus, for example, the support for the Beta distribution is the Bernoulli distribution's *p* parameter. The Bernoulli support vector contains only 2 elements, failure (0) and success (1), but the support for the prior on *p* is uncountably infinite, because there are uncountably many different possible values for *p*.

The prior distributions for the Bernoulli, the geometric and the exponential distributions are 1-dimensional because they have only 1 parameter. The support for a prior distribution on the parameters of the Normal distribution is two-dimensional, because the Normal's parameter vector has two elements (for example, its mean, μ, and sigma, σ). The support set for the prior distribution on the Normal's parameter vector is the cross product of two uncountably infinite sets: It consists of every possible combination of values for μ, which ranges from minus to plus infinity, and for σ, which ranges from 0 to +infinity. There are, of course, uncountably many combinations.

Commented [rf1]: I think variance is actually more common.

Not for my audience--crg

Prior distributions also have parameters. They are called *hyperparameters* to distinguish them from the parameters of the corresponding source distribution. For the common distributions we here consider, the number of hyperparameters is twice the number of source-distribution parameters to be estimated.

The source distribution represents uncertainty about what the value of the next datum will be; the prior/posterior distribution represents uncertainty about the value(s) of the parameter(s) of the source distribution, given the finite amount of data from which we have estimated the parameter(s).

Posterior Distributions

In a frequentist approach, one typically gathers a set of data—fills out a prespecified sample—and then computes estimates of the parameter(s) of the source distribution. One does not assume a form for the source distribution, because the central limit theorem

assures us that the sample means will be normally distributed almost regardless of what the form of the source distribution is. In a Bayesian approach, by contrast, one assumes a form for both the source distribution and the prior. The assumption about the form of the source distribution is implicit in the prior distribution, because the parameters of the source distribution are the support for the prior. (In some denotations of Bayes Rule, the dependence on the assumed form for the source is made explicit, but often it is not.)

Rather than working with samples of pre-specified size, it is not uncommon to update the posterior distribution over the parameter(s) of the source distribution datum by datum—*either* as the data come in *or* post hoc, as one considers, for example, more and more trials or more and more responses or more and more reinforcements.

The *updated* posterior distribution is often referred to as the *prior* (as in "integrating over the priors"). This is potentially confusing, as one usually thinks of the prior as the distribution before seeing any data. However, we can also think of the prior as being our belief about future data based on past data. The fact that one and the same distribution is regarded as the posterior distribution at one time—typically when it has just been updated—and as the prior distribution at another time—typically when one is about to bring in more data—takes some getting used to. However, this terminology is deeply engrained in the Bayesian approach to estimating parameters.

Consider for illustrative example the problem of estimating quickly and accurately subjects' timing coefficient of variation (CoV) from the distribution of stop latencies in the peak procedure. This distribution is known to be approximately Normal (C. R. Gallistel, King, & McDonald, 2004). Estimating the CoV requires estimating both the mean and standard deviation. For reasons to be explained when we come to conjugate priors, a good choice for the prior is the Normal-gamma distribution, which has 4 parameters. We know from extensive prior research that the mean will be positive. Although a subject may occasionally stop before the target time has elapsed, it will on average stop after that time. We also know from extensive prior research that the standard deviation will be less than half the mean. Because experimental science is a cumulative enterprise, it makes sense to take advantage of this hard-won prior knowledge. We do that by bringing it to bear on our choice of initial values for parameters of the Normal gamma. Bringing in that information can substantially reduce the amount of data required to estimate the coefficient of variation to a desired level of accuracy. Moreover, by updating the prior datum by datum, we can stop as soon as we have the desired precision in our estimate, because the updated posterior distribution on the CoV gives us a measure of the precision we have attained (the *credible interval*). Intuitively, the credible interval is the interval over which the plot of the posterior distribution is distinguishably above the x axis (aka, the support).

When using informative priors, one should bear in mind that if the data do not agree with the prior, the parameter estimates will be badly biased by the prior when there is little data. The inappropriateness of a prior will become evident if the parameter estimates after a modest amount of data diverge substantially from the mean of the initial prior distribution.

A common misunderstanding is that a prior distribution is an early version of the assumed source distribution. Purge oneself of this misconception! Repeat some large number of times: “The support for the prior is the parameter vector for the source; it is not the possible values that data may take.” The source distribution represents uncertainty about what the value of a datum may be. The prior distribution represents uncertainty about what the value(s) of the parameter(s) of the source distribution may be.

Conjugate Priors

For practical work, it is often advantageous to use a *conjugate prior*. A conjugate prior has *the* mathematical form that makes updating the prior maximally simple. It is maximally simple because the form does not change. This property is unique: One can assume whatever form for a prior one thinks makes sense; however, if one chooses a form other than the conjugate form, the posterior will no longer have the same form as the prior. Moreover, the posterior will often not be "analytic"—therefore, not one of the distribution functions made available in the standard scientific programming languages. One has to proceed numerically, which can be tricky and tedious.

Using the conjugate form for the prior has several advantages:

- The form of the posterior does not change.
- Therefore, when the prior is updated, only the values of the hyperparameters change.
- The new values are computed from the old values and from the new data by *update formulae,* which are often computationally trivial.
- The update formulae take as their arguments the previous values of the hyperparameters and some basic sample statistics (usually sums and counts).
- If one chooses any form for the prior on the Bernoulli other than the beta form, then one has to compute the source distribution, take the product between it and the prior distribution function, and compute the integral of that product over the parameters of the source distribution to obtain the normalization factor. That is intimidating, both conceptually and practically
- The conjugate prior for a given source distribution, if it exists, is unique.

In short, many practical and some purely mathematical considerations suggest using the conjugate prior. Doing so greatly simplifies the computations one has to do and it reduces the burden of defending one’s choice of a form for the prior.

Three Common Source Distributions and Their Conjugate Priors

In this primer, we deal with the three most common source distributions: the *Bernoulli*, the *exponential* and the *Normal* (aka Gaussian). Their conjugate prior distributions are the beta, the Gamma and the Normal-Gamma.

Both the source and the prior distributions may be parameterized in different ways. The different possible parameterizations can cause confusion and opportunity for error when using the distribution functions in a programming language. Make sure your programming language parameterizes a distribution in the same way you parameterize it. If it does not, use an appropriate change of variable formula. In the Supplementary Material, we list the different parameterizations of these distributions and their conjugate priors. We also provide the change-of-variable formulae.

To get started—before one brings in data—one has to assign *initial values* to the hyperparameters. We denote the initial hyperparameter vectors by θ_0 (or theta0 in code documentation). Thus, in what follows, θ without subscript refers to the parameter vector of the source; θ_0 to the initial value assumed for the prior's parameter vector, and θ_n to the parameter vector of an updated posterior. For many—but not all(!)—purposes, one wants to use a minimally informative prior, which means one wants to assign initial values that have a noticeable impact on the estimated source parameter vector only when there is very little data (e.g., 1 datum).

Often, even when one knows that one does have prior information, one wants to pretend ignorance, because ignorance is often equated with lack of bias. Also, specifying priors that actually do take into account what one already knows arouses anxiety the first few times one does it. If for whatever motive, one wants to be (or appear to be) unbiased, one should use the Jeffreys prior. It has a small—and most importantly—a readily defensible "bias."

A *Jeffreys prior* is a conjugate prior with a special and unique choice of initial value(s) for its hyperparameter(s): $\theta_{\text{beta0}} = [.5\ .5]$; $\theta_{\text{gam0}} = [.5\ 0]$; $\theta_{\text{ng0}} = [0\ 0\ {-.5}\ 0]$. Jeffreys priors are *minimally informative*. They have the further technical advantage that the parameter estimates obtained are *invariant under a change of parameters*. What that means is that, if one chose to work with a different parameterization of the source distribution —for example, with mean and variance rather than mean and precision—and if one worked with the equivalent forms for the prior distributions (the prior distribution after transformation by the change-of-variable formula), then the estimates obtained for the source distribution's parameters would agree with the estimates obtained using the alternative parameterization. It is startling and a bit disconcerting to learn that this will not be true for any choice of prior other than the Jeffreys prior! In practice, the disagreements are negligible except when there is very little data. However, we, like our subjects, are interested in the conclusions one may rationally draw when there is almost no data.

The formulae for updating the values of the hyperparameters and the custom function calls for performing these updates and plotting posterior distributions are in the Supplementary Material

Figure 3 plots the estimated source distributions and the posterior distributions on their parameter(s) for different amounts of simulated data. The left column plots the estimated Bernoulli sources and their beta-distribution posterior on the source's p parameter, given 1, 5 and 20 draws from a Bernoulli distribution whose true p value was 0.5. The initial parameter vector for the beta distribution was $\theta = [.5\ .5]$, which makes it the Jeffreys prior

on the Bernoulli. The estimated values of p and $q = 1–p$ are shown on the estimates of the source distribution. The updated values for the α and β hyperparameters (the parameters of the beta posterior) are shown on the posterior. Note that the support for the source distribution are the integers 0 and 1 ("failure" and "success"), while the support for the beta posterior is the interval from 0 to 1, the uncountably infinite number of different possible values for a Benoulli p.

The middle column of Figure 3 plots the estimates of an exponential source distribution and the gamma distribution posterior on its rate parameter, given 1, 3 and 10 draws from an exponential distribution whose true rate parameter was 0.2 responses/s. This rate corresponds to an average inter-response interval of 5 s. The estimated value for the rate parameter, λ, is shown on each plot of the estimated source distribution, along with its reciprocal, the estimated value of the mean. The initial values of the gamma hyperparameters were $\theta = [.5\ 0]$. Those initial values make the gamma distribution the Jeffreys prior on an exponential source distribution. The updated values for these hyperparameters are shown on the plots of the posterior distribution (even rows).

The right column of Figure 3 plots the estimated Normal distribution and the Normal-Gamma posterior distribution on its mean and precision (=1/var), given 1, 4 and 20 draws from a Normal distribution with a mean of 2 s and a standard deviation of .4 (therefore, a CoV of 0.2).

An *informative prior* was used to illustrate what one might do in estimating a temporal CoV. It was $\theta_0 = [0\ 0\ 4\ .37]$. It asserts that, before we have seen any real data, we have seen 4 "ghost" data— that would yield the sufficient statistics needed to estimate a variance. A variance is the mean of the squared deviations. To compute it, you need the sum of the deviations and the number of deviations that went into that sum. The 4th element in the informative θ_0 is a suggested sum of squared deviations and the 3rd element is the number of deviations on which this suggestion is notionally based.

For two reasons, the 3rd element is the one that had to be considered first in constructing this informative prior given our prior knowledge of the ballpark in which the variance should fall: i) It determines the weight given to our prior knowledge: the bigger that number, the more more informative the prior. ii) One needs that number to convert a variance into a sum of squared deviations. The starting point for the conversion was the prior knowledge that the standard deviation will probably be less than 0.3. Another way of stating that knowledge is that the variance will probably be $.3^2 \leq .09$. (Squaring the σ to get the variance is an example of a change of variable formula.) To get that variance given an n of 4.5 (the Bayesian version of 4), the sum of squared deviations has to be $4.5*.09 = .37$ (another example of applying a change of variable formula).

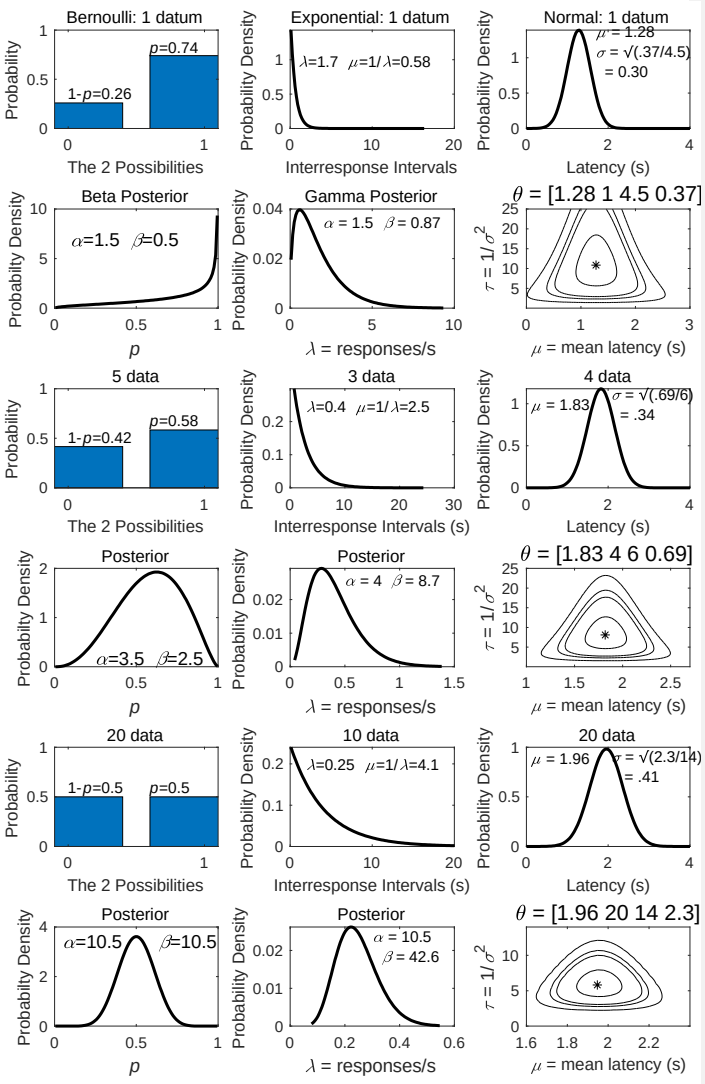

Figure 3. *Estimated source distributions (odd rows) and the corresponding conjugate posterior distributions (even rows) for the Bernoulli source (Col 1), the exponential (Col 2), and the Normal (Col 3). The estimate of the source distribution's parameter(s) is shown on each source plot. The updated hyperparameters are shown on (Cols 1 and 2) or above (Col 3) each posterior distribution. The posteriors in Col 3 are contour plots, because the posterior depends on two variables. The asterisk marks the maximum likelihood point (the summit). The contour levels are at 0.5, 0.1, 0.05 and 0.01 times the summit level. Note that the 1st element in the hyperparameter vector is the estimate of the mean. This estimate is not biased by the informative prior; it biases only the variance. Thus, the estimate of the mean given only 1 datum is the value of that datum. Without the informative prior (see text), it would not be possible to estimate the precision given only one datum. The informative prior supplies the estimate of the variance when there is but 1 datum and biases later estimates.*

The posterior on the Normal is a contour plot on a 2D support plane. The support plane contains the points that are the cross product of the plausible values for the mean and the

plausible values for the precision. The contours in a contour plot enclose the combinations that have a likelihood above some given level. They are the contours on the posterior distribution "hill," just as the contours on a topographic map are the equal-elevation contours on real hills.

Note for those who know Bayes' Rule

What is here called the *source distribution* is usually called the *likelihood.* The *likelihood function* plays a fundamental role in *Bayes' Rule*,

$$p(\theta|\boldsymbol{D}) \propto p(\boldsymbol{D}|\theta)\pi(\theta|M). \quad (1)$$

Here, $p(\theta|\boldsymbol{D})$ is the posterior distribution on the model's parameter vector given that data (D**)**, θ; $p(\boldsymbol{D}|\theta)$ is the *likelihood function,* and $\pi(\theta|M)$ is a prior distribution on the possible values for the parameter vector of an assumed stochastic model, M. The M denotes what we have called the source distribution, but with a twist: One ordinarily thinks of a source distribution—say the Bernoulli—as specifying the probabilities of various outcomes (failure or success) *given* a value for the parameter vector (e.g., given p = .5). That is, the source distribution is a probability distribution over *possible* data. The likelihood function, on the other hand, treats *actual* data—the observed outcomes—as parameters of the source distribution. Using the source distribution "backwards"—with the data taken as its parameters—gives the *likelihood* (N.B, not the *probability*) of different possible values for the source distribution's parameter. Think of M as the distribution function in a scientific programming language: When run in the forward direction, it generates the probabilities (or probability densities) for different possible values for the data. When run backwards, treating the data as parameters, it generates likelihoods for different possible values of the source distribution's parameters.

For example, assume one has observed 3 failures and 1 success when drawing from a Bernoulli distribution whose parameter, p = 0.5. The number of successes follows a binomial distribution. For this example, the source distribution for the binomial is the probability of each of the 5 possible outcomes—0, 1, 2, 3 or 4 successes— *given* that p = 0.5. The likelihood, on the other hand, is the probability of having observed 3 failures and one success for the all the values p might possibly assume. Put another way, the function $p(\boldsymbol{D}|\theta)$ is the source distribution when viewed as a function of the data, $\boldsymbol{D}$, but it is the likelihood function when viewed as a function of the parameters, θ. Unlike source distributions, likelihoods functions do not integrate to 1, which is why they are not probabilities.

Our expression for Bayes Rule (1) asserts a proportion ($\propto$), not an equality (=). That's because when the right-hand side is integrated over θ (or summed for discrete variables), it doesn't equal 1, whereas the integral of the left-hand side does (because it's a probability distribution). The factor by which the right-hand side of (1) must be rescaled is called the *normalizing factor*. It is the reciprocal of the integral. This product is sometimes called the *marginal likelihood* or the *model evidence*. For many purposes this product is all one needs. Some examples are: i) in computing a point estimate for the source parameter(s) and a credible interval on that estimate; ii) computing Bayes Factors. That is one reason why the

Commented [rf2]: If M really does go here, I'm lost.

Commented [rf3]: I couldn't really understand this, in large part because you seem to be using M to mean two things. In any case, I don't see that it adds much to the explanation directly before it. Other readers have told me this was very helpful and it seemed to help the audience in my lecture.

M DOES mean two different things. That's why the usage is confusing. When run in the forward direction, it maps possible values for data to probabilities; when run backwards, it maps data to likelihoods. These are two different functions. It's confusing to use one denotation for two different functions. The notation is not confusing for those who understand, but the language that goes with the notation is very confusing. Trust me, I was very confused for a year or two.

Commented [rf4]: This seemed like too much text for a simple idea.

Our audience does not have your talent for mathematics

normalizing factor is often omitted from the functional form of Bayes Rule and the equals sign replaced by the proportion sign. The other reason is to keep the expression as simple as possible.

Fundamentals of Information Theory

We do experiments to gain information. Intuitively, some experiments produce more information than others. The information we have gained from past experience enables us to predict/anticipate what may happen next and to infer what may have happened in the past. The information from observing outcomes enables us to infer the events and processes that produced them. This equally true of the information that non-human animals gain from their experiences in Pavlovian and instrumental conditioning experiments. It enables them to anticipate what will happen and the consequences of their actions. It also enables them to infer models of the processes and events that produce their experiences (*model-based learning*).

The study of timing behavior is the study of how brains acquire and use the information provided by objectively measurable associations (see below for how they may be measured). It cannot be distinguished from the study of associative learning, because associative learning supervenes on a temporal map (Balsam & Gallistel, 2009; Chandran & Thorwart, 2021; Honig, 1981; Taylor, Joseph, Zhaoc, & Balsam, 2014). The temporal map—a time-stamped record of past episodes—makes possible the computation of the intervals between events. That computation makes possible the inference of a predictive model.

The preceding two paragraphs presuppose we understand what information *is*. Until, 1948, one could only babble when asked to say what it was. Shannon (1948) made it a scientifically useful concept by defining it mathematically[1]. Thus, we suggest that students of timing and associative learning learn to measure the information that events provide about the form and parameters of the stochastic processes that generate those events. Bayesian parameter estimation works together with simple information-theoretic computations in a modern timing research toolkit.

Commented [rf5]: Measure seems like the wrong word, since it implies that it's a physical object light height or weight. You could replace "measure" with "compute", but I think "defining it mathematically" was his main contribution. Of course, that makes the quote problematic.

One of my main contentions is that we show how to measure the strength of an association, where an association denotes an objective fact about our experience. Most measurements require computation, particularly in their modern form. Take for example laser-based measurements of distance. Even the measure of durations requires computation because it requires differencing two times, as we discussed. Think tik and tok in Matlab; there must be equivalents in C and Python and R

To understand Shannon's definition of information, we need to understand entropy. The entropy of a probability distribution, commonly denoted by H, is given by

$$H = \sum_{i=1}^{i=n} p_i \log_b \frac{1}{p_i} \qquad (1)$$

where p_i is the probability of the i^{th} member of the support set, n is the number of elements

[1] "In physical science a first essential step in the direction of learning any subject is to find principles of numerical reckoning and methods for practically measuring some quality connected with it. When you can measure what you are speaking about, and express it in numbers, you know something about it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind; it may be the beginning of knowledge, but you have scarely, in your thoughts advanced to the stage of science." (Thomson, 1883 -p.72)

in the support set (the number of possibilities, which could be infinite) and b is the base of the logarithm. The base, b, can be any number greater than 1. In practice, it is usually e (the base of the natural logarithms) or 2; the units of entropy are *nats* in the first instance and *bits* in the second. Because entropy depends on $\log(1/p)$, it must be non-negative.

Entropy may also be considered to be a measure of uncertainty: the higher the entropy of one's distribution on some empirical variable, the more uncertain one is about the value that variable will take when next encountered. For instance, a die has higher entropy than a coin, because the chances of correctly guessing the outcome of a die role is 1/6 while the chances of correctly guessing the outcome of a coin flip is 1/2 (so you're much more likely to correctly guess the coin than the die). A biased coin has lower entropy than an unbiased one, because you're more likely to correctly guess the outcome. In the extreme case when the probability of, say, heads is 1, the entropy is zero, and you're sure of the outcome before the coin is flipped.

Entropy of a continuous distribution.
Most of the time, continuous functions and distributions can be discretized, and if the discretization is fine enough the quantities one cares about don't change. For example, we could replace a probability distribution $p(x)$ with its discretized version, in which the probability that a variable lies between x and $x+dx$ is $p(x)\,dx$. In the limit of small dx, the discrete distribution still sums to 1 (or very close to 1), and we can still do statistical inference. Moreover, as dx goes to zero, those operations become increasingly accurate. However, one thing we can't do is compute entropy. It's easy to see why: entropy is a measure of uncertainty, and as dx becomes small, we become increasingly uncertain which interval our variable lies in. As dx goes to zero, we become completely uncertain, and the entropy goes to infinity. It does not make sense to refer to the probability attached to a point in the support for a continuous distribution, because there are uncountably many of them and there is no way to even refer to all but a negligible number of them (the countable ones).

When confronted with infinities, the best thing to do is simply throw them away. This is what early information theorists did: they defined the differential entropy analogously to Equation (1), but for continuous distributions,

$$H[p(x)] = \int dx\, p(x) \log_b (1/p(x)). \qquad (2)$$

The problem with throwing away infinities is that it's hard to do rigorously. And, in fact, the numerical value of differential entropy varies with the units attached to the data. Changing the units can give it any value one likes. It also changes under a nonlinear change of variables, often by a very large amount – a point we'll return to shortly.

Fortunately, we're almost always interested in differences in entropy, which are better behaved. A particularly important difference is the *mutual information*, I, between two variables, say x and y, defined as

$$I(x,y) = H[p(x)] - \int dy\, p(y)\, H[p(x|y)] \qquad (3)$$

where the notation $p(x|y)$ means the probability distribution over x conditioned on y. Although it's not immediately obvious, $I(x,y)$ is independent of the units with which one measures x and y. It's even invariant under a nonlinear change of variables. The above definition of mutual information applies to discrete distributions as well (simply replace the integral over y by a sum). It even applies to mixed discrete and continuous distributions.

Mutual information has a natural interpretation of obvious psychological and neuroscientific importance: it's an upper limit on the average reduction in uncertainty about x one can get from observing y—and vice versa, because $I(x,y) = I(y,x)$. When y is a direct measure of x, but with error bars, then information is approximately equal to the entropy of $p(x)$ with x expressed in the units equal to the error bars. For instance, if observing y told us the value of x to within 1 cm, then $I(x,y) \approx H[p(x)]$ if x is measured in cm. However, this expression is valid only if $H[p(x)]$ is large; it breaks down if $H[p(x)]$ is small, and it breaks especially badly if $H[p(x)]$ is negative (since mutual information cannot be negative).

Unfortunately, not all differences in entropy are so well behaved. If we have two distributions $p_1(x)$ and $p_2(x)$, then $H[p_1(x)] - H[p_2(x)]$ doesn't depend on the units of x. However, it does change under a nonlinear change of variables. For instance, the differential entropy of an exponential distribution is the natural log of its rate parameter. However, if we switch to log units, the entropy becomes independent of the rate parameter. This is relevant, for instance, for time, because log units often make more sense than linear units. It's why we have units of time that increase exponentially (seconds, minutes, hours, days, etc.), and it's consistent with the Weber's law for time, which tells us that errors in estimating time scale linearly with time, and so are constant in log time.

On the other hand, the consequences of changing the variable from time to log time have unacceptable intuitive/interpretive psychological and neuroscientific consequences, because brains do not treat either time or number logarithmically. They do not treat the difference between 1second and 2 seconds as equivalent to the difference between 1 hour and 2 hours, even though log2(7200) – log2(3600) = log2(2) – log2(1) = 1 (Brannon, Wusthoff, Gallistel, & Gibbon, 2001; Gibbon & Church, 1981). Nor do human subjects think that their uncertainty is the same when told that a bus will arrive in the next few minutes versus in the next few hours. In information theory, entropy and uncertainty are often treated as two sides of the same coin, a duality that goes back to Shannon (1948). We want to preserve these intuitive meanings when bringing information theory to bear on psychological and neuroscientific issues.

Available information, the other side of the entropy coin. Consider a random variable, x, for example, the interval between two successive randomly scheduled rewards. We denote by p(X) a probability distribution on x. When the rewards are randomly scheduled, p(X) is an exponential distribution. Consider another random variable, y, and a probability distribution on it, denoted by p(Y). When we have decided on appropriate units, the *available information* about x is the entropy of p(X); it measures the *uncertainty* about the

true value of *x*. When learning a value for *y* removes all uncertainty about the value of *x*, then *y* communicates all the available information. The amount of uncertainty that has been removed depends on the precision with which we have measured *x*. In scientific notation (where '1201' is denoted by 1.201e3), the precision is (or at least should be) reflected in the number of digits to the left of e. When there is only 1 decimal digit, the implied precision is 1 part in 10; when there are 4 decimal digits, as in the example given, the implied precision is 1 part in 10,000, a precision neuro-biologically made measurements rarely attain.

A message cannot reduce a subject's uncertainty by more than the entropy in the subject's distribution on the variable in question. While from a mathematical perspective, the support vector for a continuous distribution is uncountably infinite, the same cannot be true for a subject's representation of a continuous distribution, because neurobiologically realized representations of quantities must have limited precision. Limited precision makes the number of elements in a support vector countable. All physically realizable systems for symbolizing quantities and carrying out computations on those symbols have limited precision. Good systems also take account of the limits on precision that derive from imprecision in the measurements that map from the quantities to the symbols that represent them. The extent to which the results of a computation can be trusted depend on the quality of the measurements that delivered the numbers that went into the computation. (The garbage-in-garbage-out principle.)

Shannon's Coding Theorem

Shannon's (1948) coding theorem proves that, in the maximally efficient code for data coming from some distribution, the length of the code for a given datum is proportional to $\log(1/p)$, where p is the probability of that datum—the relative frequency with which it has to be encoded. His theorem entails that when one is coding the data from a distribution with a given form (e.g., the Bernoulli or the exponential or the Normal) *and* with a given parameter vector, θ, then, to maximally economize on the amount of memory used to store the data, one must adjust the coding scheme so that the length of the code words—for example, the number of bits in a binary vector that encodes interval durations—grows as the log of the inverse of the probability that the brain will need to encode a given interval. Intuitively, rare events must be assigned long code words and frequent events short code words— and the function relating relative frequency to code length must be logarithmic. Shannon's coding theorem is the foundation of modern communication technology; it tells us how to make maximally efficient use of physical resources such as memory and signal bandwidth. Without Shannon's insights, there would be no Zoom and no Netflix.

A consequence of Shannon's coding theorem is that, when using a well-constructed coding scheme, the lengths of the words used to store data are a physical realization of the (log of) the probability vector in the distribution that models the data distribution. Therefore, the relative frequency of a datum may be determined from the relative length of the "word" that encodes it. The length of the word that encodes a datum may be thought of as the height of a bar in a histogram with a logarithmic y axis. The width of that bar is the range of x values (e.g., experienced intervals) to which that probability is assigned. The intervals

that map to that bar are treated as "essentially the same" for the purpose for which the data are currently represented. For what 'the length of the word' might mean physically, think bit pattern or nucleotide sequence.

The Kullback-Leibler Divergence

The Kullback Leibler divergence *of* a distribution, P, *from* a distribution, Q, is denoted $\mathrm{Dkl}(P||Q)$. It gives the average cost (usually in bits or nats) of encoding a datum from the P distribution using a code optimized for the Q distribution. In other words, the cost of erroneously assuming that the two distributions are one and the same. The prepositions 'of' and 'from' are stressed because the divergence is not symmetric, that is, $\mathrm{Dkl}(P||Q) \neq \mathrm{Dkl}(Q||P)$.

Some information theorists consider the Kullback-Leibler divergence to be a more basic or foundational information-theoretic measure than entropy. Unlike entropy, it is well defined for both continuous and discrete distributions and invariant under change of variable.

When n data have come from a distribution, Y, that diverges from a distribution, X, by $\mathrm{Dkl}(Y||X)$, the cumulative number of memory bits that have been wasted encoding the y's on the assumption they were x's is nDkl. We call this the *cumulative coding cost*. We show in an Appendix that it maps to the probability that two distributions with the same form do not differ, given the data: We show that the nDkl is distributed $\Gamma(n_p/2,1)$, where Γ denotes the gamma distribution and n_p is the number of parameters (e.g., 1 for the Bernoulli and exponential, 2 for the Normal). Thus, the nDkl is a simple information-theoretic measure of whether there is a significant difference between the distributions.

Measuring Association and Contingency

Events are temporally associated to the extent that the location of the next event may be predicted from knowledge of the location of the preceding event, and vice versa. A random distribution of event times maximizes uncertainty about where in time the next event and the preceding event may be found. In information-theoretic terms, it is the *maximum entropy distribution* (Jaynes, 1957, 2003). The maximum entropy principle is an information-theoretic formulation of Occam's razor: assume as little as possible. The maximum entropy distribution for events distributed in time is the exponential.

One way to think about the exponential distribution is that in any (infinitesimal) time bin dt, the probability that a food pellet appears is rate $\times$ dt. This means events are completely randomly distributed in time; in other words, they are not *self-associated*. Knowing, for instance, the most recent t_x does not alter an observer's uncertainty about where in time the next t_x may be encountered nor where in the temporal map the preceding t_x may be found. Any other distribution induces some degree of self-association; that is, the location of the next point can to some extent be predicted from the location of the preceding point and vice versa. This gives the exponential distribution a very counter-intuitive property: Suppose we repeatedly drop a pointer onto the time line at randomly chosen points in time,

and we compute the intervals looking forward in time from the pointer to the next t_x and also backward in time to the most recent t_x—the *prospective* and *retrospective* intervals). They both have entropy $H(X)$, where X is the exponential distribution.

Consider now a stream of y events, [y_1, y_2, ...] occurring at times t_1^y, t_2^y .... We want a measure of the maximum possible uncertainty about t_{n+1}^y given t_n^y. It follows from the counter-intuitive facts about the exponential distribution that the maximum possible uncertainty is measured by the entropy of an empirical exponential distribution with $\mu = n/D$, where n is the number of y events so far observed and D is the duration of observation. The more predictable t_{n+1}^y becomes, the more this entropy is reduced. When $t_{n+1}^y - t_n^y$ is a constant, the distribution of the intervals between successive y events has no entropy, t_{n+1}^y is completely predictable when given t_n^y, and the y events are maximally self-associated.

Consider now a second stream of x events occurring at times t_1^x, t_2^x ..... We want a measure of the extent to which the x events are associated with the y events, a measure of how *pre*dictable the next t^x is when given a t^y. We also want a measure of how *retro*dictable the preceding t^y is when given a t^x.

The predictability of the next t^x given a t^y is maximized when $H(X|Y) = 0$ and minimized when $H(X|Y) = H(X)$. Similarly, the retrodictability of the preceding t^y when given a t^x is maximized when $H(Y|X) = 0$ and minimized when $H(Y|X) = H(Y)$. The *maximization* in both cases (prediction and retrodiction) occurs only when t^x and t^y always coincide, that is only when $\forall n t_n^x \equiv t_n^y$ and the distribution of these coincident events is exponential. The minimization of predictability (maximization of uncertainty) occurs when both distributions are independent. In that case, the mutual information is 0, because $H(X|Y) = \mathrm{H(X)} - H(X|Y) = 0 = H(Y) - H(Y|X)$. Finally, in the conditions of applicability we here consider, $H(Y|X) \leq \mathrm{H(X)} \geq 0$.

A measure related to the mutual information between X and Y may therefore be constructed as follows:

- Let C be an *unconditional distribution* of intervals, with rate parameter $\lambda|$C. [In the examples considered, these intervals will be the inter-reward or inter-shock intervals when the subject is in a test chamber in which a transient CS, such as a noise or light, creates mutual exclusive and exhaustive periods denoted by CS and ~CS. That is why we denote the unconditional distribution by C: one can think that it means either **C**hamber or **C**ontext. In operant conditioning (reinforcement learning), the contextual distribution is the distribution of the intervals between rewards (or, more generally, between act *outcomes*.]
- Let Y be the *conditional distribution* of intervals, with rate parameter $\lambda|$Y. [In excitatory Pavlovian conditioning, this is the distribution of waits for reward signaled by CS onsets. In inhibitory Pavlovian conditioning and in trace conditioning, it is the distribution of waits for rewards signaled by CS offsets. In operant conditioning, the *retrospective conditional distribution* is the distribution of intervals looking back from the rewards to the most recent response. There is also a

prospective conditional distribution in operant conditioning, but its definition differs depending on the protocol (VI, FI, FR, VR, etc).]

- The contextual and conditional distributions are always chosen such that $\lambda|\mathrm{C} \leq \lambda|\mathrm{Y}$, the contextual rate is less than or equal to the conditional rate.
- The sufficient statistics from a sample of wait intervals are the number of waits and the duration over which they were observed. Under the maximum entropy principle, this is equivalent to treating the distributions as exponential.
- Entropies are computed using the formula for the differential entropy of the exponential, $1-\ln(\lambda)$, where λ is the rate parameter $(=1/\mu)$
- The unit of time is chosen so that $\lambda|\mathrm{C} \leq \lambda|\mathrm{Y} < 1$ This makes the formula for the differential entropy of the exponential unproblematically applicable under the stipulated restrictions.
-

The proposed measure of association is:

$$\Delta H|\mathrm{Y\&C} = \big(1-\ln(\lambda|\mathrm{C})\big) - \big(1-\ln(\lambda|\mathrm{Y})\big) = \ln\big(\ln(\lambda|\mathrm{Y})\big) - \ln\big(\ln(\lambda|\mathrm{C})\big) = \ln\frac{\lambda|\mathrm{Y}}{\lambda|\mathrm{C}} \qquad (4)$$

given that $[\lambda|\mathrm{C} \leq \lambda|\mathrm{Y} \leq 1]$, and otherwise undefined.

And the contingency, denoted $\mathcal{C}(X;Y)$, is (under the same restrictions):

$$\mathcal{C}(X;Y) = \frac{\Delta H|\mathrm{Y\&C}}{1-\ln(1/k)} \qquad (5)$$

In words—using a well-known example— Equation (4) says that the association between the CS and the US in excitatory Pavlovian conditioning is measured by the reduction in uncertainty about the waits for reward following the onset of a CS. Equation (5) says that the contingency is that reduction normalized by the *available information*, the amount of information that a CS could convey. the denominator. The procedure for estimating k is explained in the next section.

The Time-Scale Invariance of Association and the Estimation of k

The proposed measure of association, Equation (4), does not measure the strength of a hypothetical construct in the mind or brain, such as a connection weight, or the strength of a Hebbian synapse or the value attributed to a reward; rather, it measures an observable fact about the temporal distribution of events. The rate ratio in (4) is unitless. If associative learning depends on what ΔH measures, it must be time-scale invariant, because informativeness is time-scale invariant. There is extensive evidence that Pavlovian conditioning is time-scale invariant (C. R. Gallistel & Gibbon, 2000). The evidence first emerged in a meta-analysis of trials to acquisition in pigeon autoshaping done by Gibbon and Balsam (1981).

Pigeon autoshaping is a Pavlovian protocol in which an illuminated key takes the role of the bell (the CS) and pecking that key takes the role of salivation (the conditioned response). It

was studied intensively by many labs in the 1970s because it proved to be a more efficient way of training pigeons to peck keys than the shaping recommend by Skinner (1938).

Until the discovery of autoshaping, it had been assumed that teaching a pigeon to peck a key was the paradigmatic example of reinforcement learning (aka operant conditioning). The difference between Pavlovian conditioning and operant conditioning is that in operant conditioning, the learning depends on the subject's observing the consequences of behavior; it seems to be driven by a retrospective association, a look back in time. In Pavlovian conditioning, the behavior is irrelevant to the learning process. The only role of behavior in Pavlovian conditioning is to reveal to the experimenter whether or not the subject has learned (perceived) the association. The learning in Pavlovian conditioning appears to be driven by a prospective association, a look forward in time.

Balsam and Gallistel (2009) suggest that the rate ratio in (4) be called a protocol's *informativeness*, because it determines the amount of information a subject may gain from a CS. Gibbon and Balsam (1981) called it the C/T ratio. C stood for μ_C the **C**ycle duration, that is, the US-US interval), and T stood for μ_{CS}. In the delay-conditioning protocols they analyzed, the CS had a fixed duration and the US coincided with CS offset. Thus, the wait for the US at CS onset was the **T**rial duration.

In an autoshaping protocol, the key is illuminated every now and then (at more or less random intervals) for a fixed duration, at the end of which the food (US) is delivered. It is delivered regardless of anything the subject may have done. Different labs used different trial durations (denoted T, the duration of the CS) and different different US-US intervals. They denoted the US-US interval by C for the cycle duration. It is the sum of CS duration and the average ITI, that is, the average interval between the termination of the previous trial and the onset of the next CS.

The now widely accepted operating definition of rate of learning—the reciprocal of USs to acquisition—was then little attended to. It was often not reported for individual subjects, as is now best practice. However, Gibbon and Balsam obtained the raw data from twelve different labs, which enabled them to compute for each bird, the trial at which it satisfied an acquisition criterion (one or more pecks on three out of four successive trials).

They discovered a surprising regularity (Figure 4): The data are well described by a simple regression equation: $n(\lambda|CS/\lambda|C\text{-}1) \geq k$, where $k = 294 \pm 28$. The regression model applies over the full range of learning rates, from 0 (infinite USs to acquisition) to 1 (acquisition following the first US), a span of almost 3 orders of magnitude on both axes. It accounts for 75% of the variance, with no evidence of systematic deviation, as evidenced by the out-of-sample circle in Figure 4, a datum that was not in the data to which the model was fit (the asterisks).

The success of this regression model is a theory killer. We are not aware of a formalized theory of associative learning that can produce it. Moreover, it has only one free parameter, *k*. That parameter has a data-based interpretation; it is the $\lambda|CS/\lambda|C$ that produces one-trial learning. In Equation (5), we suggested that the entropy of an exponential with rate parameter $1/k$ be considered the available information, the upper limit on how much information a Pavlovian CS can provide.

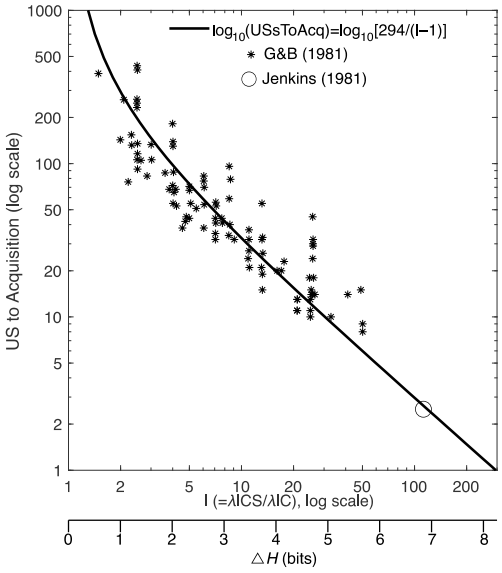

Figure 4. *Median USs to the acquisition of a conditioned response in pigeon autoshaping protocols, as a function of the informativeness (I) of the protocol, on double-logarithmic coordinates. The asterisks are the data plotted in Figure 7.11 on p. 245 of Gibbon & Balsam (1981). The regression line was fit to those data. The open circle is an out-of-sample datum from Group No F in Table 8.5 on p. 245 of Jenkins et al (1981). The second x axis, shows the relation between informativeness and its logarithm, ΔH, the suggested measure of associative strength. The 95% confidence interval on k is 265 to 322 (0.2 bits).*

The quantity on which the rate of learning depends, $\lambda|CS/\lambda|C - 1$ is the percent increase in the rate of reward to be expected when the CS is present ($\lambda|CS$), relative to the rate expected simply from being in the context in which the CS occurs ($\lambda|C$). That makes intuitive sense on what might be called the make-hay-while-the sun-shines principle, where the pecking the key is understood as foraging behavior (making hay).

The regression equation further implies that subjects begin to respond to the CS only when the cumulative excess USs exceeds *k*. One would like to understand what determines this decision criterion. Figure 5 may give a hint. The Dkl a measure of the strength of the evidence that $\lambda|CS$ differs from $\lambda|C$ (see Appendix). It also gives the amount of memory that may be saved by recognizing this difference by using $\lambda|CS$ rather than $\lambda|C$ to encode waits experienced during CSs. Figure 5 shows that the value for *k* in the regression equation is the $\lambda|CS/\lambda|C$ such that the Dkl corresponds to a *p* value $< .002$ against the hypothesis that $\lambda|CS = \lambda|C$. It also shows that adopting the $\lambda|CS$ estimate when encoding the waits for USs during CSs will save $\Delta H|CS) = 8$ bits (one byte) of memory per datum encoded.

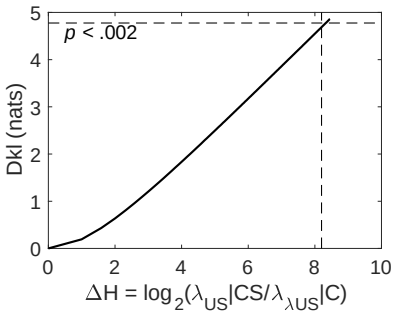

Figure 5. *The Dkl plotted against* $\Delta H|CS = log_2(\lambda|CS/\lambda|C)$*. The vertical dashed line is the* $\Delta H|CS$ *when the protocoled association produces one-trial learning. The horizontal dashed line is the probability of receiving this amount of information from a spurious association.*

One might expect the decision to respond under circumstances that do not produce one-trial learning would be based on the strength of the evidence accrued over the pre-decision trials, that is, on nDkl, the *cumulative* coding cost. That is not the case: Near the other end of the regression, when $\lambda|CS/\lambda|C = 1.5$, the Dkl is only .07 bits, indicating that the evidence for an association on any given CS-US pairing is very weak. Weak associations may be spurious and not assumed to exist until frequently observed.

However, acquisition does not occur until the 500th trial, when the nDkl is > 36. That value for the nDkl indicates astronomical evidence for the association. If acquisition were based on the accumulated evidence for a weak association, it would have occurred much sooner, because the nDkl would have exceeded 8 bits after $8/.072 = 111$ USs. Thus, the rationale for basing the decision to respond on the cumulative excess USs rather than on the cumulative evidence for an association remains to be clearly understood. However, Figure 5 suggests that the choice of a criterial cumulative excess may depend on the excess that, if experienced on a single occasion, would be evidence for a maximum-strength association. Put another way, one way of understanding k is that $1–\ln(1/k)$ is the available information, which is why it appears as the normalizing factor in our proposed definition of contingency, Equation (5). Given the magnitude of k, $1–\ln(1/k)$ is essentially $\ln(k)$, which is $1.44*\ln(k)$ in bits.

Time-Scale Invariance and Contingency in Reinforcement Learning

The conclusion that Pavlovian conditioning depends on the time-scale-invariant $\Delta H|CS$, which measures the association between CS and US timing, poses the question whether the same is true in reinforcement learning (aka operant conditioning). Previous work in the information-theoretic framework (C. R. Gallistel, Craig, A., Shahan, T.A., 2019) implicates the importance of two different associations in reinforcement learning—the prospective association, which is the extent to which a response predicts reward, and the retrospective association, which is the extent to which a reward retrodicts a response.

If the processes that detect these associations are time-scale invariant, then an arbitrarily long hang-fire latency between an act and an outcome—between response and reinforcement— should be no obstacle to the maximally rapid learning of an operant response. It should be learnable after only one or two rewarded responses when there is a strong contingency between the response and the expected wait for the next reward, that

is, when either: 1) a response communicates substantial information about when to expect reward (prospective contingency) or 2) the reward communicates substantial information about the recency of a response (retrospective contingency).

Prospective and retrospective contingencies are not the same because the distribution of intervals looking back from rewarding outcome (O) to the most recent act (A) may be very different from the distribution of intervals looking forward from A's to the next O (C. R. Gallistel, Craig, A., Shahan, T.A., 2019). When pigeons peck a key on variable interval schedules of reward, every reward is preceded at a very short, fixed interval by the peck that triggers its delivery. Thus, the entropy of *reward-conditional* distribution of the *retrospective* intervals to a response, $H(A|O)$, is 0, while the unconditional (marginal) distribution has substantial entropy, because inter-peck intervals are approximately exponentially distributed (C. R. Gallistel, Craig, A., Shahan, T.A., 2019). Therefore, adapting Equation (4) to the present case:

$$\Delta H\left(A;\overleftarrow{O}\right) = H(A) - H(A|\overleftarrow{O}) = H(A) - 0 = H(A) \qquad (6)$$

In words, when the retrospective conditional entropy, $H(A|\overleftarrow{O})$, is 0, knowledge of the time at which the outcome occurred reduces to 0 the uncertainty about the recency of the act because act and outcome are coincident to within the accuracy of measurement. Thus, the outcome communicates all of the available information about the recency of the act that produced it. Therefore, the *retrospective contingency*, $\Delta H\left(A;\overleftarrow{O}\right)/H(A)$, is 1.

On the other hand, pigeons pecking on a variable interval schedules of reward peck at a much higher rate than the rate of reward. Although the delivery of the reward is triggered by a peck, the effect-less pecks between the rewards and the reward-triggering peck are so numerous that the distribution of intervals looking forward from pecks to the next reward—the distribution of A–O intervals—is practically indistinguishable from the distribution of O–O intervals (C. R. Gallistel, Craig, A., Shahan, T.A., 2019). In that case,

$$\Delta H\left(O;\vec{A}\right) = H(A) - H(O|\vec{A}) = H(A) - H(A) = 0 \qquad (7)$$

so the *prospective contingency*, $\Delta H\left(O;\vec{A}\right)/H(A)$ is 0.

Degrading the retrospective contingency

Lengthening the hang-fire interval between a reward-triggering act and the delivery of the triggered outcome (reward delivery) allows (reward-irrelevant) acts to intrude into the hang-fire intervals. The intrusion of these acts adds entropy to the retrospective conditional distribution, $A|\overleftarrow{O}$. The longer one makes the hang-fire interval, the greater this entropy becomes; hence, the lower the perceivable retrospective contingency becomes. Gallistel et al (2019) found that subjects responding on VI schedules with lengthened hang-fire intervals reduced their rate of response so as to maintain a critical amount of $\Delta H\left(A;\overleftarrow{O}\right)$.

This result, together with some little known previous results on instrumental learning with 30s delays of reward (Lattal & Gleeson, 1990), led Gallistel et al (2019) to conjecture that the computation that solves the assignment of credit problem in reinforcement learning is time-scale invariant.

The assignment-of-credit problem in reinforcement learning poses the question, What did I do that made that happen? How brains solve this problem is a central concern of computational neuroscientists working on Reinforcement Learning (Dayan & Niv, 2008; Gershman, Norman, & Niv, 2015; R.S. Sutton, 1984; R. S. Sutton & Barto, 1998). If the credit-assignment process is time-scale invariant, then the interval between an act and the outcome it triggers can be arbitrarily long, provided that the naive response rate is low enough so that the retrospective intervals between initial rewards and the responses that trigger them are much shorter than initial estimates of reward-reward intervals.

A recent experiment in Shahan's lab, now being written up, tested this conjecture with the following simple protocol: Naive rats were given four half-hour long sessions of magazine training during which they learned that the 3s illumination of the feeding hopper signaled the release of a food pellet. This hopper training was followed by an hour-long session of context extinction, during which no pellets dropped into the hopper and there were no hopper illuminations. The 10 subjects were then divided into an experimental group and a group of yoked controls ($n = 5$ in both groups).

Both groups were returned to their test boxes, in which a lever was now extended. For subjects in the experimental group, pressing it triggered the drop of a pellet into the hopper (and illuminated it for 3s coincident with the drop)—but only after a hang-fire delay of 2 minutes. Presses made during the hang-fire delay had no consequences.

When a subject in the yoked control group pressed the lever, it had no consequences. However, they experienced the same pellet releases and hopper illuminations as the subject in the experimental group to which they were yoked.

To the best of Shahan's knowledge, a 2-minute delay is longer than any delay of reinforcement ever tested in an operant experiment. Ever since Skinner's seminal work (Ferster & Skinner, 1957; Skinner, 1938), operant conditioners have supposed that more or less "immediate" reinforcement of responses was critical. They have, however, remained non-committal about the definition of 'immediate'. This same supposition found in most contemporary reinforcement-learning models: the reinforcement is assumed to be delivered at the termination of the state in which the causal act is made (Gershman et al., 2015; Yael Niv, 2019; Y. Niv, Daw, & Dayan, 2005).

Positing an "I just made a response state" that endures for 2 minutes seems a stretch. During two minutes, awake rats generate many different responses. They may also make the same response many times. Thus, this experiment poses in particularly stark form the question of how brains solve the assignment of credit problem in reinforcement learning. How do they learn what works and doesn't work? How fast do they learn it? What are the crucial experiential variables that determine the answers to these questions? And, perhaps

most importantly, what is the representation of their experience that enables them to solve the problem? Does reinforcement learning also supervene on a temporal map, the learning of which makes possible the computation of prospective and retrospective interval durations?

Estimating Prospective and Retrospective Contingencies After the First Few Rewards

By Equation (6), the prospective $\Delta H\left(O;\vec{A}\right)$ is $\log_2\left(\lambda_{\mathrm{O}|\vec{\mathrm{A}}}/\lambda_{\mathrm{O}}\right)$. $\lambda_{\mathrm{O}|\vec{\mathrm{A}}}$ is the (estimated) rate of the outcome following the act [=1/(average wait for that outcome after making that act)], while λ_{O}. is the marginal (unconditional) rate of reward [1/(average inter-reward interval)]. Thus, to measure the prospective and retrospective associations, we must estimate $\lambda_{\mathrm{O}|\vec{\mathrm{A}}}$ in order to estimate the entropy of the waits for reward conditioned on the subject's having made a response and we must estimate λ_{O} in order to estimate the entropy of the waits for reward (the entropy of the unconditional or marginal distribution).

The response-conditional rate of reward,$\lambda_{\mathrm{O}|\vec{\mathrm{A}}}$, cannot be less than $1/2$ minutes given the protocol, because the wait for a reward after making a response is never greater than 2 minutes. The *average* wait will, however be shorter than 2 minutes if a subject makes further responses during the wait triggered by an initial response. These intruding responses do not trigger rewards, but they do reduce the average wait between a response and the next reward. Four of the 5 experimental subjects made additional responses during the 2-minute delay following their first response. The closer these additional responses came to the reward, the shorter the average A–>O interval. Thus, it was generally less 2 minutes, particularly early in training. Thus, $\lambda_{\mathrm{O}|\vec{\mathrm{A}}} \geq 0.5/\mathrm{min}$ for the experimental subjects. For their yoked controls, on the other hand, the average wait for a reward was the average wait from randomly chosen points in time. If the distribution of O-O (reward-reward) intervals is approximately exponential, then the average wait for a reward from a randomly chosen points in time is equal to the average O-O interval (that is, the contextual inter-reward interval).

How to estimate the marginal distribution (the unconditional waits for reward) is ambiguous—for us, and probably for the rats as well. They spent 60+ minutes in the test box prior to the first reward. If one takes the 60 minutes with no reward during context extinction into account, then $\lambda_{\mathrm{O}} < 1/60\ \mathrm{min} = 0.0167\ \mathrm{min}^{-1}$ after the first reward. In that case, $\Delta H\left(O;\vec{A}\right) = 4.9$ bits.

The ambiguity about the relevant intervals for computing the marginal entropy arises from the fact that the lever was not present during context extinction. The rats may have taken its presence as a change in context, because the new context enabled an action that was not possible in the preceding context.

If rats regarded the box with a lever as a new context, then their estimate of the contextual rate would have been based only on the latency of the first reward in the first session with

the lever present. That latency ranged from 2.8 minutes to 13.6 minutes, yielding unconditional rates of reward of $1/2.8 = 0.36$ to $1/13.6 = .074$ rewards/min.

The initial values for the response-conditional rates of reward in this context depend on the initial pattern of responding. For a subject that makes only 1 response before reward delivery, the initial response-conditional rate of reward is 0.5 min^{-1}. In that case, the prospective ΔH would range from $\log_2(.5/.36) = 0.47$ bits to $\log_2(.5/.074) = 2.8$ bits. Suppose, however, that a subject makes a first response, waits 110 seconds and then makes 9 more responses in the 10 seconds prior to reward delivery. The average wait for reward following a response is then mean([1 2 3 4 5 6 7 8 9 120]) = 16.5s, for a response-conditional rate of reward of $60/16.5 = 3.6$ min^{-1}. This rate is much higher than any of the unconditional rates of reward in the lever context. On the other hand, suppose the subject responded at 60 responses/minute during the entire 2 minutes between its first response and the reward triggered by that first response. In that case, the average interval from a response to a reward would be 1 minute, and a response-conditional rate of reward = 1/min. This highlights the fundamental difference between the associations that drive reinforcement learning and the associations that drive Pavlovian learning: In Pavlovian protocols, the associations between CS and reward do not depend on the subject's behavior, while in operant protocols, both the prospective and the retrospective associations do depend on the subject's behavior.

Figure 6a plots the prospective $\Delta H\left(O; \vec{A}\right)$ over the first 10 rewards for the first pair of yoked subjects when the marginal entropy, $H(A)$, was estimated from reward-by-reward Bayesian estimates of $\lambda_{O|\vec{A}}$. and reward by reward estimates of λ_O using only the intervals observed in the context where the lever was present. In this pair, the $\triangle H$ was already a measurable quantity (equal to almost 1 bit) after the first response made by both the experimental subject and its yoked control. (They happened to make their first presses at almost the same elapsed time in the session.) This objective aspect of the experimental subject's experience was already strong (greater than 5 bits) after the experimental subject's 2nd response; whereas, for the yoked control, it dropped to near 0 after its 3rd response. To learn what these two subjects did, readers will have to wait for the publication.

Figure 6b plots the retrospective $\triangle H$ over the first 10 rewards for the same paired subjects. It, too, became almost immediately very strong for the experimental subject and 0 for the yoked control. Thus, there is a readily measurable objective aspect of each subject's experience that could explain an immediate difference in their behavior after a single experience in which there was a 2-minute separation between the causal action and the outcome it produced.

Estimating the strength of the evidence

The reward-by-reward estimates of the prospective and retrospective associations in Figure 6—that is, the $\triangle H$'s on the y axes—were computed by Bayesian estimation of the rate parameters. The estimate after 1 reward was based on one datum; the estimate after 2 rewards on two data, and so on. There is, of course, great uncertainty about the accuracy of

these estimates, and that uncertainty propagates to the estimates of the prospective and retrospective associations. One obviously wants measures of the strength of the evidence for them. That's where the nDkl and the posterior distributions on the rate estimates come into play. The posterior distributions quantify the uncertainty about the parameter estimates and the nDkl measures the strength of the evidence provided by two different rate estimates.

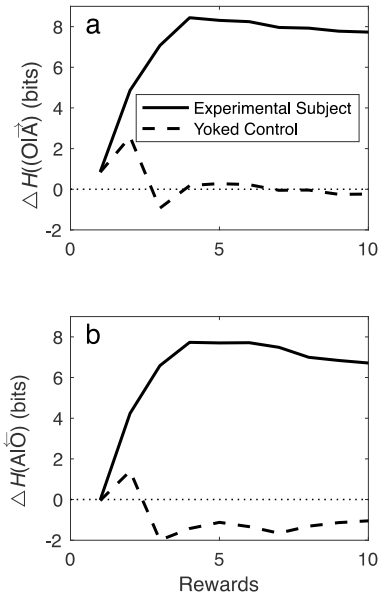

Figure 6. **a.** *The prospective change in entropy (reduction in uncertainty),* $\Delta H\left(O;\vec{A}\right)$*, as a function of number of rewards, for one experimental subject and its yoked control. Only the first 10 rewards are shown.* **b.** *The retrospective change in entropy,* $\Delta H\left(A;\overleftarrow{O}\right)$ *for the same pair*

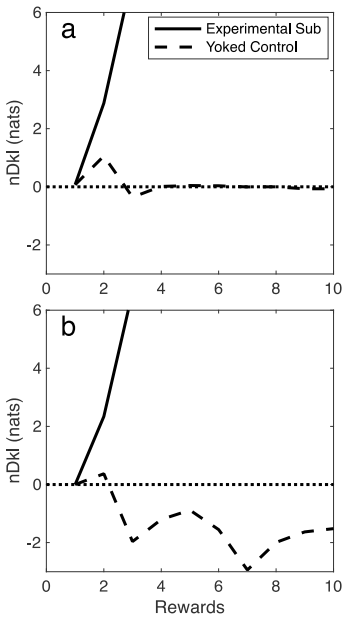

Figure 7. **a.** *The cumulative coding cost of assuming no prospective association between a response and the wait for reward, as a function of the number of rewards, for both the experimental and the yoked subject. When this cost exceeds 1.92 nats, the evidence for the association is significant at beyond the .05 level. When it exceeds 5.4 nats, it is significant at beyond the .001 level.* **b**. *The cumulative coding cost of assuming no retrospective association between a reward and the recency of the last response, as a function of the number of rewards, for both the experimental and the yoked subject. For an explanation of how the nDkl acquired its negative sign, see text.*

The Dkl is gives the cost of encoding a wait drawn from the conditional distribution on the assumption that it comes from the unconditional (marginal) distribution. When that assumption is true, the Dkl decreases as a scalar function of the sample n—a manifestation of the law of large numbers—with the consequence that the *cumulative* coding cost, nDkl, has the same distribution regardless of the n. We show in the Appendix that nDkl is distributed $\Gamma(.5,1)$ when the distributions are exponential. Because the formula for the Dkl is itself extremely simple—$\text{Dkl(in nats)} = \ln(\lambda_P) - \ln(\lambda_Q) + \lambda_Q/\lambda_P - 1$—the nDkl provides an exceptionally simple measure of the evidence for a difference between two exponential distributions. As a bonus, it has a neuroscientific interpretation; it gives the number of bits of memory storage space that may be saved by recoding the already observed conditional waits. Figure 7 plots the nDkl's for the prospective and retrospective ΔH's against the number of reward.

In Figure 7b, the cumulative coding cost for the yoked control is moderately negative. A divergence, like an entropy, cannot be negative. However, to facilitate graphic interpretation, we have added to the custom functions that compute and plot the nDkl an option that allows the user to give the nDkl the sign of the difference between the conditional and the marginal rate estimates. When there are few data, spurious associations may appear giving rise to smallish nDkl's that are in the wrong direction. In looking at the graphs, one does not want to confuse the effects of small-sample error, which diminish as the sample grows, with real effects. The real effects always grow as the sample grows. Our adding sign to indicate divergences opposite to those expected (and observed!) in the long run explains the initially negative cumulative coding costs in the retrospective cumulative coding cost for the yoked control in Figure 7b. The asymptotic cumulative cost was within about ± 1 of 0 (data not shown), as it must be when events are independently distributed in time.

Measuring the Strength of the Evidence for Differences in Probability

To illustrate the application of Bayesian parameter estimation and the cumulative coding cost to Bernoulli probabilities, we draw on data from a not-yet published experiment conducted by Basak Akdogan in Peter Balsam's lab. Her experiment used tone durations as the discriminative stimuli (S△), in a two-lever operant choice procedure, with mice as subjects. In operant conditioning, an S△ is a signal that indicates which of two possible acts will produce a reward. It is presented just before two levers appear, enabling a choice of actions. In initial training, the S△ was a tone lasting either 2s or 6s. For the subject whose data we analyze here, the choice of the left lever was rewarded following the 2s tone and the choice of the right lever was rewarded following the 6s tone.

The subject was pretrained until it chose the correct levers above chance following both tone durations. When the subject had been responding at asymptote for 600 trials, the discriminative stimuli changed: The 2s tone no longer occurred; it was replaced by an 18s tone. The correct response to this novel duration was the left lever—the lever that had previously been correct given the 2s tone, that is, the shorter of the two initial S△'s. The 6s tone continued to occur on 50% of the trials. The correct response on those trials remained what it had always been: press the right lever.

The last 2s tone occurred on Trial 601; On Trial 602 the tone lasted the already familiar 6s. On Trial 603, it lasted for a never-previously-experienced 18s. The subjects had no way of knowing what the consequences of pressing either lever might now be. They also had no way of knowing how frequently to expect the 18s duration. If its occurrence signaled a change in the state of the world. In this new state, there was no way to know how frequently to expect the other two durations (2s and 6s) nor what the reinforcement contingencies might be.

In this and most experiments of a similar nature, the first statistical issue is estimating a subject's pre-switch probability of choosing the correct lever following a given S△ and the uncertainty about what that value is. A more challenging issue is to determine whether pre-change choice performance is/was stable.

We estimate the pre-change p_{correct} using the Jeffreys prior, which is the beta distribution with initial hyperparameters $\theta_{\text{beta}} = [.5\ .5]$. When updated by the number of correct choices, n_s, and failures to choose correctly, n_f, over the last 300 pre-change 6s trials, the (hyper)parameters of the beta prior/posterior are $\theta_{\text{beta}} = [n_s{+}.5\ n_f{+}.5]$. Figure 8 plots the posterior distribution on the pre-change probability of a correct choice following a 6s tone. This distribution represents the uncertainty about the estimate of the subject's probability of a correct choice.

We can compute *critical intervals* on our estimate of p from θ_{beta}, using the inverse function in the suite of functions that scientific programming languages provide for distributions in common use (see Code for each Figure in the Supplementary Material). *Critical intervals* are the Bayesian version of *confidence intervals*, but they have a less convoluted interpretation: The ratio of the area under the probability distribution within a critical interval to the area that falls outside that interval is the odds that the value of the estimated parameter lies within the critical interval, given the data. Using the beta inverse function, we find that only 1% of the area under the curve in Figure 8 lies below 0.85 and only 1% lies above .93; thus, the odds are 50:1 in favor of the conclusion that the subject's pre-change probability of choosing the right lever was in the interval between [.85 .93].

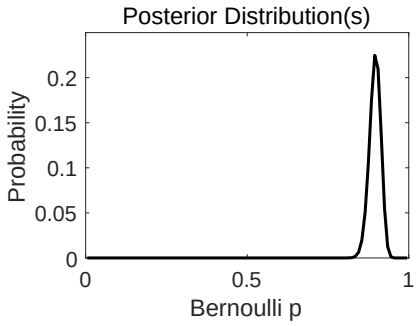

Figure 8. *The posterior beta distribution on the estimate of the Bernoulli probability, p, of a correct choice following the 6s prior to the change in the discriminative stimuli.*

Checking on the stability of a parameter estimate

An often-vexing methodological issue is the criterion for when a subject has attained asymptotic performance or, at least, a stable level of performance. The *n*Dkl statistic can help.

To check on the stability of the pre-change lever-choice probability in this mouse, we call a custom function that compares an evolving *p* value to a reference value and computes the *n*Dkl statistic to identify stretches of trials where there is strong evidence of a deviation from the reference distribution:

[CmPdif,nDkl,PnDkl,pt] = BernCCCchange(D1,theta1,theta0,.9,true);

D1 is the binary vector of successful (rewarded) choices of the right lever during the pre-change era; theta1 is the updated parameter vector for the beta posterior distribution as of the final (300th) pre-change trial; theta0 is the initial vector of hyperparameters for the beta prior; and the optional 'true' (the 5th input) tells the function to plot the figure (see Figure 9). The fourth input argument, .9; is the complement (1–α) of a "significance" level (α) for the *n*Dkl statistic. Including it among the input arguments causes the function to return NaN when the number of data and the reference *p* are together such that a "significant" *n*Dkl is impossible. For example, when the α is 0.05 and there are fewer than 5 data, an *n*Dkl significant at α is impossible, because the probability of getting 4 heads in the first 4 flips of a fair coin is .0625

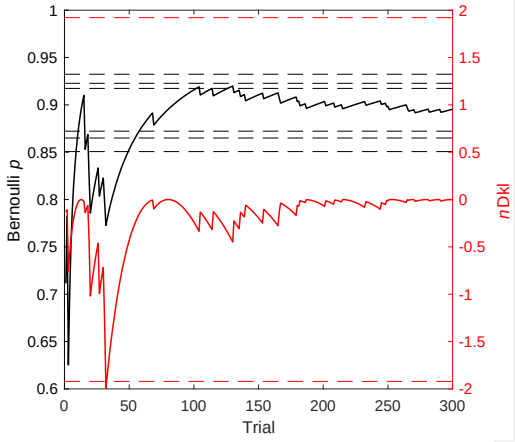

Figure 9. *The trial-by-trial estimate of the probability of choosing the right lever as a function of the pre-change sequence of trials (black curve, plotted against left axis). The thin black dashed lines give the critical levels for this estimate, given the complete data set. The red curve is the signed nDkl statistic, plotted against the right axis. The thin red horizontal dashed lines (bottom and top of panel) represent alpha levels of .05 on this statistic.*

Sign was added to the plot of the nDkl red curve in Figure 9 to indicate the direction of the divergence. The subject's estimated probability of pressing the right lever following a tone of 6s duration was lower than the lower limit of the critical intervals on the terminal estimate during Trials 1-8 (when a significant departure of $\hat{p}$ from a reference value of 0.9 was impossible) and then again from Trials 19-49, when a significant departure was entirely possible (black curve in Figure 9). However, the signed *n*Dkl reached moderate significance (dashed red horizontal at bottom of plot) on only a single trial (Trial 31, $p<.05$). The fact that this trend did not continue, and nDkl turned back toward zero,

indicates that it is a statistical fluctuation that implies no departure from a stable choice probability of 0.9. In general, the longer one continues to flip a fair coin, the more certain it becomes that one will observe atypical sequences that seems to imply the coin was not fair over that stretch of flips. Thus, *brief* excursions in *n*Dkl beyond essentially arbitrary alpha levels, such as the one that valleys at Trial 31 in Figure 9, should be ignored. The fluctuations in the nDkl already observed between trials 0 and 150 fall well within those expected on the null hypothesis, so this experiment could have gone on to the next phase much sooner had these analytic methods been used on line while the experiment was running.

The convergence of the red curve plotting *n*Dkl to close to 0 as the number of trials approaches 300 is a peculiarity of this data set. The distribution of the *n*Dkl under the null hypothesis is independent of *n*; it does not become narrower as *n* grows larger. The fact that it's close to zero in this plot just means that the subject's probability of pressing the right level was within about 1 part in 300 of 0.9. If the experiment went on longer, the nDkl would eventually explore values within about ± 1 of 0.

Measuring the Growing Strength of Stochastic Stimuli

From an information-theoretic perspective, conditioning protocols are stochastic stimuli unfolding in time because the amount of information available to the subject about the contingencies in the protocol increases as the protocol persists.

The subject's behavior is a stochastic stimulus for the experimenter and/or the data analyst: As we observe more behavior, the evidence for (or against) a contingency between the SΔ and a subject's choices appears and grows stronger. The cumulative coding cost allows us to compare the growth of the *objective* evidence of reinforcing contingencies observed by a subject—the strength of the stimulus as a function of time—to the strength of the *behavioral* evidence for contingency perception, as a function of time.

When the 2s duration ceased to occur and a novel 18s duration began to occur, the mouse was confronted with a novel discriminative stimulus (a tone lasting 18s). A question of central interest was the rapidity with which the mouse would adapt its behavior to the contingency between this new stimulus and reinforcement.

A consideration of fundamental importance in the analysis of *instrumental* behavior is that the rate at which the subject acquires information about the true state of affairs depends on its behavior. It depends on what Reinforcement Learning theorists call *exploratory* behavior, and we call *information-gathering* behavior. This is what distinguishes instrumental conditioning (aka operant conditioning) from Pavlovian conditioning.

One of the many interesting aspects of Akdogan's experiment is that it pits the ideal *observer* against the ideal *agent*. The ideal observer is often taken to be the observer that performs perfect statistical inference given the data. However, the performance of perfect statistical inference presupposes that the observer has the correct model. More

importantly, this conception of the ideal observer implicitly assumes that the observer's behavior has no effect on the data it has seen (and will see!).

The ideal agent, by contrast, behaves so as to maximize its *return*, the amount of some desired outcome attained per unit time invested in acting. A properly informed agent is one that has gained the knowledge necessary to act optimally a given the assumed loss function.

The Growth of Behavior-Independent Probability Estimates

Among the things the subject does not know after the first 18s tone is the relative frequencies with it it will again hear the three DS durations it has so far encountered. Figure 10 plots Bayesian estimates of the post-change probabilities of hearing each tone duration as a function of the number of post-change trials.

The middle panel of Figure 10 plots the ideal observer's estimate of the probability of a tone lasting 6s, on the assumption that the first occurrence of the 18s tone leads this ideal observer to wonder whether all bets are off. In this computation, the observer's uncertainty about that is captured by putting a weakly informative prior of $\theta_0 = [5\ 5]$ on the new probability of a 6s tone. (Note the contrast to the uninformative prior in which $\theta_0 = [0.5\ 0.5]$). The red plot in that panel is the cumulative coding cost of assuming that the new probability of a 6s tone (when estimated using a weakly informative prior) is the same as the old one. The nDkl is stable and low, giving no suggestion that this probability has changed.

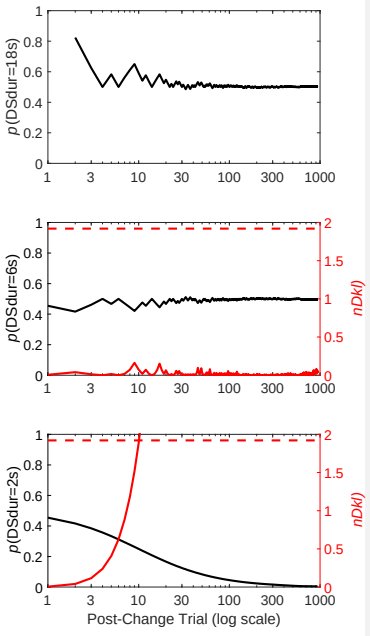

Figure 10. Top: *Bayesian estimate of the probability of an 18s duration tone as a function of the number of trials, counting from its first occurrence.* **Middle:** *Bayesian estimate of the probability of the 6s ⊿S given a weak presumption that it continues to be 0.5, (black curve, plotted against the left axis) and the (unsigned) nDkl statistic for its divergence from the pre-change probability (red curve, plotted against the right axis). The thin red dashed line at top of plot is the .05 alpha level on the nDkl.* **Bottom:** *Bayesian estimate of the probability of the 2s ⊿S (black, left axis) and nDkl (red, right axis). The odds that it has diminished exceed 20:1 after the 11th post-change trial.*

By contrast, the black curve in the bottom panel of Figure 10 plots the Bayesian observer's estimate of the probability of a 2s tone, on the same assumption, while the red curve plots the cumulative coding cost of

making that assumption. The odds against the no-change assumption are 20:1 after the 11th trial.

Because of the informative prior, the Bayesian estimate of the new probability is 0.23 after 11 successive trials during which a 2s duration has not occurred. The increasing odds against the no-change hypothesis give reason to abandon the informative prior. If one replaces it with the uninformative Jeffreys prior, $\theta_0 = [0.5\ 0.5]$, the odds against the assumption that the new probability is the same as the old are better than 20:1 after the 5th trial and the estimate of the new probability is 0.08. With the new improved (uninformative) prior, the odds against the no-change hypothesis are then 1,000:1 after the 11th post-change trial. A rational observer would change her prior, because, when assessing stochastic stimuli, the future is informative about the best representation of the past.

Commented [P6]: I'm guessing you incorporated the prior into the Dkl. You shouldn't; it was derived for maximum likelihood estimation (see text on page 32).

Yes, we need to find someplace to put this in; however, I would like to add that practice so far suggests that the impact of the .5 on the nDkl is minimal

In sum, the results in Figure 10 tell a Bayesian observer that by the 11th post-change trial, the probability of an 18s tone is approximately 0.5, the probability of a 6s tone remains approximately 0.5, and the probability of a 2s tone is trending toward 0.

Tracking the Change in the Behavioral Probabilities

In behaviorist models of choice, subjects do not learn probabilities; rather they form habits (Hull, 1930). This is called model-free learning in contemporary Reinforcement Learning theories. For the mouse whose data are here featured, the habit of choosing the right lever following a tone of 6s duration was rewarded on every 6s trial both before and after the substitution of 18s tones for the 2s tones. Its reaction to this change shows that choosing the right lever following a 6s tone was not a habit; its behavior depended on the arithmetic relation between the durations.

In reaction to the appearance of 18s tones and the disappearance of 2s tones, the mouse reduced its probability of pressing the right lever following the 6s tones, even though pressing the right lever continued to be unfailingly reinforced (Figure 11, top). The reduced probability of pressing the right lever and the correspondingly increased probability of pressing the left lever following 6s tones became evident on the 14th post-change trial, which was the 7th 6s trial following the first occurrence of an 18s tone. The behavioral change is indicated by the downward inflection in the black curve and the corresponding sharp upward inflection in the red curve in Figure 11. The odds confirming the existence of this change permanently exceeded 20:1 after the next 15 6s trials (red curve, Figure 11). At that point, the Bayesian estimate of the probability of choosing the right lever had dropped below the lower .01 boundary of the 98% critical interval on the Bayesian estimate of the pre-change probability of this choice (black curve, Figure 11).

This enduring reaction to the 6s stimulus reduced the subject's probability of reinforcement on 6s trials, hence its overall rate of reinforcement. The violation of the predictions of the model-free" habit" hypothesis was even more striking in the other 7 mice in this protocol. Several of them reduced their probability of pressing the correct lever well below chance.

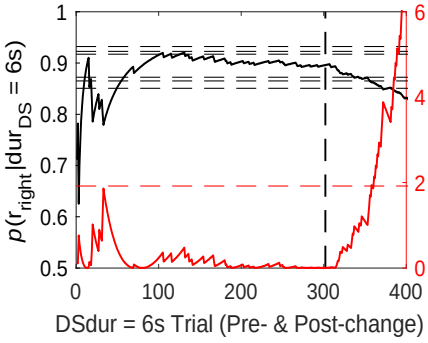

Figure 11. *The conditional probability of sampling (pressing) the right lever as a function of all the trials (pre- and post-change) on which the ∆S was 6s (black curve plotted against left axis) and the cumulative cost of assuming the post-change probability equals the pre-change (red curve, plotted against right axis). The thin, black, dashed, horizontal lines indicate upper and lower limits on critical intervals for the pre-change estimate (intervals containing .8, .9 and .98 of the probability mass). The odds against the null hypothesis are 20:1 above the thin, red, horizontal dashed line. The thin vertical dashed black line indicates the first occurrence of the 18s tone.*

The sustained and substantial reduction in the post-change probability of choosing the right lever following a 6s tone can be understood on the assumption that subjects place a high value on information in a changing world. When things change, it pays to behave so as to learn the new contingencies, because knowing them is a pre-condition for optimal behavior. When the 18s tone supplants the 2s tone, for all the mouse knows, pressing the left lever following a 6s tone may sometimes yield a bigger reward than that yielded by pressing the right lever. Continuing to press the left lever only very rarely on 6s trials will retard the forming of an estimate of what those two probabilities might be—the probable size of a possibly bigger reward and the probability of producing it. Thus, the rationality of a subject's post-change behavior can only be judged when we know the value it places on the information to be gained about the variety of consequences that *might* follow from pressing the left lever on 6s trials relative to the value it places on maintaining the previously experienced rate of reward on those trials.

Measuring Contingency Detection Behaviorally and Photometrically

Kalmbach, et al (Kalmbach et al., 2021 under review) measured mesolimbic dopamine activity photometrically in mice that had previously learned to press a lever for food reward. The photometric monitoring of dopamine activity began when these mice first began to hear tones that lasted 80s, during which lever presses did not produce rewards. In other words, the already learned contingency between pressing a lever and obtaining food was now contingent on the absence of the tone (a second order contingency).

A CS subdivides the context in which it occurs into mutually exclusive and exhaustive CS intervals and ~CS intervals. aka intertrial intervals, or ITIs for short. When calculating associative strength, the conditional distribution must always be the distribution whose rate of reward is higher than the contextual rate of reward. Thus, the conditional

distribution in this protocol is the distribution of US-US intervals during the ITIs. Its rate parameter is λ(US|~CS), the informativeness is λ(US|~CS)/λ(US|C), and Δ*H* is the log of that rate ratio.

Figure 12 plots the trial-by-trial rate estimates and the nDkl for two subjects. In Figure 12a, the subject began to respond at a higher rate during the ITIs than during the tones only after 300 trials. In Figure 12b, the subject consistently responded at a higher rate during the ITIs after the 8th trial. This 37-fold difference in the rate of learning is an extreme example of the noisiness commonly seen in this statistic (trials to acquisition).

To delimit the training interval within which the conditioned behavior or neurobiological activity appeared, we extracted two measures from these plots: 1) The trial after which the evidence for a CS-ITI difference in the behavior or neurobiological activity) permanently exceeded an evidentiary criterion. 2) The trial after which the estimated response rate during the ITIs permanently exceeded the estimated response rate during the CSs. This latter trial may be regarded as the trial after which the conditioned behavior appeared, while the former is the trial at which the evidence that it had appeared became decisive. Because the strength of the evidence for a change grows as more data come in, the evidence for it often becomes decisive only after the change is apparent in retrospect. These two trials—the trial after which conditioned response appeared and the trial after which the evidence for it was decisive—are marked, respectively, by a vertical dotted red line and by a vertical dash-dot red line in Figure 12.

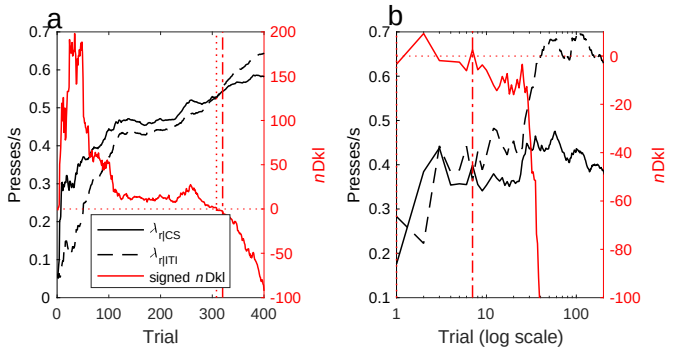

Figure 12. *Trial-by-trial estimates of the response rates during the CSs and the ITIs (black solid and dashed curves, left axis) and the signed cumulative cost of assuming that the CS rate is the same as the ITI rate (red curve, right axis). Vertical dotted red lines indicate the trial after which ($\lambda_{r|CS} - \lambda_{r|ITI}$) is enduringly negative. Vertical dash-dot red lines are the trials after which the signed nDkl remained less that 3.3 nats (the p<.01 level).* **a**. *The rate estimates (black plots) cross at Trial 308, where the vertical dotted red line is; the vertical red dash-dot red line is at Trial 320. The x axis is linear.* **b**. *The black curves cross for the last time at Trial 8. This is also the trial after which the signed nDkl is permanently less than 3.3 nats (p<.01). Therefore, the vertical red dash-dot line superposes on the vertical*

Commented [P7]: I think in panel b the vertical lines are in the wrong place: the dashed one should be moved to where the dot-dashed one is, and the dot-dashed one should be shifted to the right of that.

The value of nDkl on Trial 9 is <3.3 & ditto for Trial 10 (albeit just barely), so Trial 8 is the trial after which the nDkl is always <3.3

red dotted line. The x axis has been logged to better reveal what happened over the first 10 trials.

Figure 13 plots the signed cumulative coding cost of assuming that the CS rates are the same as the ITI rates for the 8 subjects in the negative contingency ("inhibitory") protocol (top two rows) and the 4 subjects in the truly random control (no contingency). For the 4 subjects in the non-contingent condition (bottom row of Figure 13), the nDKL was positive throughout training. Note also that these nDkl's did not continue to climb, unlike the nDkl's for the negative contingency subjects, which maintained or often increased their downward slope as training continued. The slope of the nDkl is proportionate to the difference in the rate estimates. When the slope is 0, so is the difference in the rate estimates.

Applying the nDkl to the Photometric Data on DA activity.

Abby Kalmbach recorded DA activity photometrically on most of the training sessions. Technical problems sometimes prevented her obtaining a signal on some sessions, particularly with the first few subjects. In the course of training, a marked drop in the mean signal appeared in the negative contingency subjects—the subjects in which the onset of the CS signaled a decrease in the rate of reward to below the contextual rate and its offset signaled and increase to above that rate. A striking feature of the drop was a negative spike during the first 1.5 seconds of each CS and a positive spike during the 1.5 s following the termination of the CS.

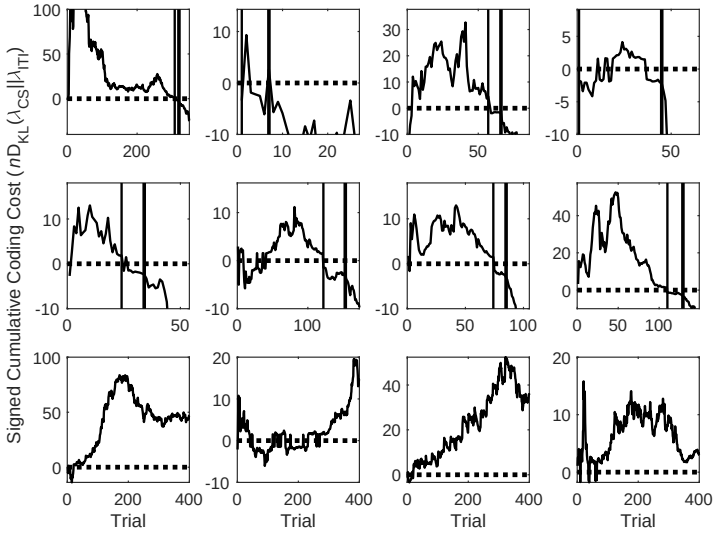

Figure 13. *The signed cumulative coding cost of assuming that the rate of pressing during the CSs is the same as the rate during the ITIs for all 12 subjects. The thinner vertical line*

indicates the trial at which the response appeared; the thick vertical line, the trial at which the evidence for it became decisive.

To measure trial-by-trial the development of these photometric spikes, we constructed templates for them by averaging those same 1.5s segments across the last 200 training trials, when the spikes were well developed. We then correlated these templates with the corresponding segments in the individual traces from the early trials. The trial-by-trial correlation coefficients were approximately normally distributed. We updated trial by trial the Normal-Gamma posterior distribution on the mean and posterior of this source distribution —the Normal distribution of the correlation coefficients.

We did not, however, use the Jeffreys values for the θ_0 of the Normal-Gamma (the Jeffreys θ_0 = <0 0 -.5 0>). We are interested in the mean value of the correlations, not their variance. The variance is what is called a *nuisance parameter*. The variance of a Normal distribution, hence its precision, which is the reciprocal of the variance, can assume any positive value. However, because these data are correlations, we have analytic prior knowledge of the variance: The variance of a distribution of correlations cannot be greater than 1. Generally speaking, it will be substantially less than 1. We also had confirmatory empirical prior knowledge: Across subjects and regardless of the protocol (negatively contingent or truly random), the variance in the correlations was approximately 0.22.

Given this analytic and empirical prior knowledge, we used θ_0 = <0 0 4 0.9>. This prior implicitly assumes that we had already seen 4 correlations (3rd element of the parameter vector) and that the sum of their squared deviations from the mean correlation was .9 (4th element). This prior biased the variance estimate toward what we knew a posteriori must be about the right value, thereby heightening the sensitivity of the nDkl. The nDkl in the Gaussian case depends on the (pooled) variance estimate as well as on the difference between the means. We did not bias the estimate of the mean, which was the parameter of interest.

In this analysis, the value for the mean of the Q distribution (the reference distribution) is 0 (the null hypothesis = no correlation). This is an analytic fact; there is no uncertainty about it; hence, there is no posterior distribution on the mean of Q. Also, P and Q are assumed to have the same variance, the variance estimated from the (gently biased) data. We therefore computed the trial-by-trial *n*Dkl's with and without integrating over the posterior distribution on the parameters of *P*. The results were essentially the same; the resulting estimates of the trial at which the nDkl began to rise and the estimates of the trial at which the odds against the null became permanently greater than 100:1 did not disagree by more than 3 trials in any subject. For two of the 5 small disagreements in the estimates of *n*Dkl, the estimate obtained by integrating over the posterior distributions on the parameters of the Normal distribution was greater than the estimate obtained using the maximum likelihood estimates for these source parameters.

Figure 14 plots the photometric nDkl's (right two columns) alongside the (negatively signed) behavioral ones (left column). For most subjects, decisive evidence (indicated by

blue verticals) for a negative photometric spike at CS onset and a positive spike at CS offset appeared sooner than decisive behavioral evidence for the detection of the negative contingency between the CS and reward delivery. However, in one subject, decisive behavioral evidence appeared very quickly and well before decisive photometric evidence (see row 2 in Figure 14). In all the subjects, the behavioral evidence rapidly got very much stronger than the photometric evidence, because the behavioral "signal" (the magnitude of the difference in response rates) got stronger soon after evidence for it became decisive. The photometric signals also tended to strengthen, leading to the moderate upward concavity seen in the nDkl's in the right two columns. The strengthening of the photometric signals was, however, less pronounced than the strengthening in the behavioral signals.

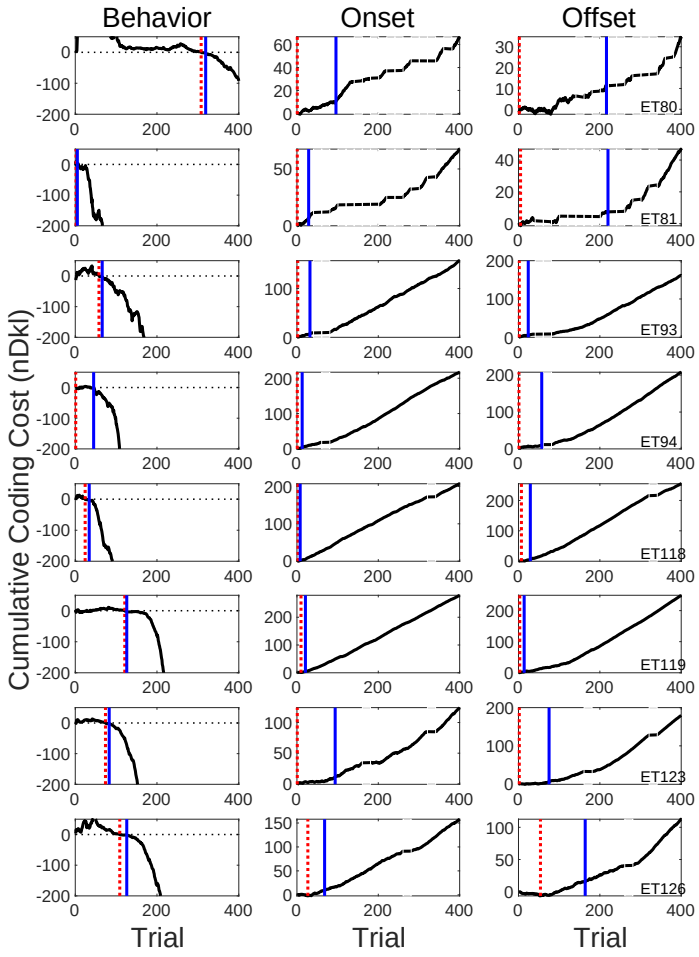

Figure 14. *The signed nDkl plots for the behavioral "signal" alongside the (unsigned) photometric nDkl's for the onset and offset spikes in the dopaminergic photometry signal. Blue verticals mark the trials where evidence becomes decisive; dotted red verticals mark the trial where it first appears. Grey verticals in the photometric columns indicate sessions where photometry signal could not be obtained. In the control subjects, the photometric nDkl's were similar to the behavioral ones (bottom row of Figure 13), in that there was little evidence for the spikes at CS onset and offset in the photometric signals from the control subjects.*

Conclusions

A temporal map of past experience enables the replay of episodes and the recovery of associative structure (Gupta, van der Meer, Touretzky, & Redish, 2010; Mattar & Daw, 2018; Ólafsdóttir, Bush, & Barry, 2018; Panoz-Brown et al., 2018; van de Ven, Jäckels, & De Weerd, 2022; Zentall, 2019). Information-theoretic tools enable the quantification of associative structure using ΔH, which is the entropy of the marginal (unconditional) distribution minus the entropy of the conditional (unconditional) distribution—computed on the assumption that both distributions are exponential.

Bayesian parameter estimation enables us to estimate the strength of these associations after the first US in Pavlovian protocols and after the first the first reinforcement in operant protocols. The nDkl (cumulative coding cost) measures the strength of the evidence for the association. The entropy difference, ΔH, is the information-theoretic analog of a correlation coefficient, while nDkl is the information-theoretic analog of its statistical significance.

The nDkl measure might prove relevant to the search for the engram (Langille & Gallistel, 2020; Poo et al., 2016), because it gives the amount of memory a brain can save by recoding the temporal map in memory using a stochastic model that takes into account the observed temporal associations. The mnemonic benefits from recoding previously stored data in the light of an improved stochastic model provide a computational rationale for consolidation and reconsolidation, which appear to be fundamental aspects of memory management (McKenzie & Eichenbaum, 2011).

Adopting new stochastic models to conserve memory resources improves a brain's ability to anticipate future rewards and punishments and to recognize the causal effects of the behavior it generates. A model that better explains the data already seen better predicts the data not yet seen, when model complexity is properly accounted for (Grünwald, 2007). These considerations suggest that information theory may prove relevant to discovering the neurobiological processes that construct the temporal map and that do the computations that lead to anticipatory behavior in Pavlovian conditioning and to operant behavior in reinforcement learning.

Whether consolidation and reconsolidation are manifestations of memory saving based on the recognition of stochastic structure proves to be true or not, these tools enable us to measure on a reward-by-reward or punishment-by-punishment basis the strength of the evidence that a subject's on-going experience provides about the contingencies we create

when we define an experimental protocol. By enabling us to measure the evolving strength of the evidence for associative structure, these tools put the study of timed behavior and associative learning on the same conceptual footings as the study of sensory processing and perception, fields where Bayesian inference and information theory now play fundamental roles (Brainard, 2009; Chater, Tenenbaum, & Yuille, 2006; Feldman, 2016, 2021; Froyen, Feldman, & Singh, 2015; Ganguli & Simoncelli, 2016; Hiratani & Latham, 2020; Maloney, 2003; Panzeri, Harvey, Piasini, Latham, & Fellin, 2016; Simoncelli & Olshausen, 2001; Stocker & Simoncelli, 2008). Using the same tools, we can measure simultaneously: i) the strength of the stochastic stimulus, ii) the strength of the evidence for it, iii) the strength of the behavioral and neurobiological changes induced by the perception of the association, and iv) the strength of the evidence for these changes.

In 1967, Rescorla pointed out that Pavlovian conditioning depended on temporal contingencies, not temporal pairing (Rescorla, 1967). He further pointed out that contingencies were determined by how events were distributed in time. He confessed, however, that he did not have a way of computing contingency. That problem has now been solved, not only for Pavlovian conditioning, but also for operant conditioning.

Contingency may be defined as the $\Delta H(\mathrm{X}|\mathrm{Y})/(1-\ln(1/k))$. This definition simplifies to $\Delta H(\mathrm{X}|\mathrm{Y})/\ln(k)$ when k is large. X denotes the marginal distribution and Y the conditional distribution: $\Delta H = \log\left(\lambda_{x|y}/\lambda_x\right)$, where the lambdas are the rate parameters of distributions assumed to be exponential. The temporal units attached to the rate estimates are such that both $\ln(\lambda)$s>0. $1-\ln(1/k)$) is the available information, the maximum amount of information that a $\Delta H(\mathrm{X}|\mathrm{Y})$ could convey.

In pigeon autoshaping, $k = 294$ [$\mathrm{CI}_{.95}$=266 322] (see Figure 4).. A similarly large value of k has also been obtained in as yet unpublished experiments on inhibitory Pavlovian conditioning with rat subjects in the Balsam lab ($k_{\mathrm{inhib}} = 260$, $\mathrm{CI}_{.95} = [216\ 303]$). A still larger k (approximately 800) may be estimated from Figure 10 in Gallistel and Gibbon (2000), which gives C/T results for rabbit eyeblink conditioning. Given these large values for k, the expression for a Pavlovian contingency simplifies to $\Delta H(\mathrm{X}|\mathrm{Y})/\ln(k)$, which ranges between 5.7 and 6.7 nats (8.2 to 9.6 bits) across diverse species (pigeons, rats and rabbits) and with both excitatory and inhibitory protocols.

When the unit of ΔH (associative strength) is bits, raising 2 to the power of ΔH gives the factor by which the presence of a CS in Pavlovian excitatory conditioning may reduce a subject's uncertainty about the wait for the next reward or punishment. In inhibitory conditioning, it gives the factor by which ~CS intervals may reduce that uncertainty. In both cases, the reduction is relative to the context in which the CS and the ~CS occur. In operant conditioning (reinforcement learning), raising 2 to the power of the prospective ΔH gives the extent to which making a response can reduce the subject's uncertainty about the wait for the next reinforcement, while raising to 2 to the power of the retrospective ΔH gives the extent to which the occurrence of a reward reduces a subject's uncertainty about where in its temporal map the most recent response occurred. This latter reduction may be taken as evidence for causality.

In this approach to associative learning, an association is not a conductive connection in a brain (a connection weight or a Hebbian synapse). Nor is it a subjective value placed on reward or punishment. It is a measurable fact about the distribution of events in time. The computations that enable the perception of this fact presuppose a temporal map, a time-stamped record of events. Its temporal map enables a brain to look back in time to compute the intervals and the rate parameters of assumed-to-be exponential distributions.

Our use of the entropy difference as a measure of temporal association is related to a more general approach to defining clusters information-theoretically (Slonim, Gurinder, Tracik, & Bialek, 2005). Events are temporally associated when they cluster in time. When they do so, knowledge of the location of one event in the cluster provides information about where the other events may be found (van de Ven et al., 2022) and evidence for some underlying causal process that explains the cluster. Clustering is a time-scale invariant phenomenon, because it is defined by the entropy within a cluster relative to the entropy of the marginal distributions, and entropy itself is time-scale invariant.

Balsam, P. D., & Gallistel, C. R. (2009). Temporal maps and informativeness in associative learning. *Trends in Neurosciences*, *32*(2), 73-78. doi:http://dx.doi.org/10.1016/j.tins.2008.10.004

Brainard, D. H. (2009). Bayesian approaches to color vision. In M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences*, (pp. 395-408). Cambridge, MA,: MIT Press,.

Brannon, E. M., Wusthoff, C. J., Gallistel, C. R., & Gibbon, J. (2001). Numerical subtraction in the pigeon: Evidence for a linear subjective number scale. *Psychological Science*, *12*(3), 238-243.

Chandran, M., & Thorwart, A. (2021). Time in Associative Learning: A Review on Temporal Maps. *Frontiers in Human Neuroscience*, *15*(167). doi:10.3389/fnhum.2021.617943

Chater, N., Tenenbaum, J. B., & Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Sciences*, *10*(7), 287-291.

Dayan, P., & Niv, Y. (2008). Reinforcement learning: The good, the bad and the ugly. *Current Opinion in Neurobiology*, *18*(2), 185-196.

Feldman, J. (2016). The simplicity principle in perception and cognition. *WIREs Cogn Sci*, *7*, 330-340. doi:10.1002/wcs.1406

Feldman, J. (2021). Mutual information and category perception. *Psychological Science*, *32*(8), 1298-1310.

Ferster, C. B., & Skinner, B. F. (1957). *Schedules of reinforcement*. East Norwalk, CT: Appleton-Century-Crofts.

Froyen, V., Feldman, J., & Singh, M. (2015). Bayesian hierarchical grouping: Perceptual grouping as mixture estimation. *Psychological Review*, *122*, 575-597.

Gallistel, C. R. (2021). Robert Rescorla: Time, Information and Contingency. *Revista de historia de la psicología*, *42*(1), 7-21.

Gallistel, C. R., Craig, A., Shahan, T.A. (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. doi:10.1037/rev0000163

Gallistel, C. R., & Gibbon, J. (2000). Time, rate, and conditioning. *Psychological Review*, *107*(2), 289-344. doi:doi.org/10.1037/0033-295X.107.2.289

Gallistel, C. R., King, A., & McDonald, R. J. (2004). Sources of Variability and Systematic Error in Mouse Timing Behavior. *Journal of Experimental Psychology: Animal Behavior Processes*, *30*(1), 3-16.

Ganguli, D., & Simoncelli, E. P. (2016). Neural and perceptual signatures of efficient sensory coding. 1–24. Retrieved from http://arxiv.org/abs/1603.00058

Gershman, S. J., Norman, K. A., & Niv, Y. (2015). Discovering latent causes in reinforcement learning. *Current Opinion in Behavioral Science*, *5*, 43:50.

Gibbon, J., & Balsam, P. D. (1981). Spreading associations in time. In C. M. Locurto, H. S. Terrace, & J. Gibbon (Eds.), *Autoshaping and conditioning theory* (pp. 219-253). New York: Academic.

Gibbon, J., & Church, R. M. (1981). Time left: linear versus logarithmic subjective time. *Journal of Experimental Psychology: Animal Behavior Processes*, *7*(2), 87-107.

Grünwald, P. (2007). *The minimum description length principle*. Cambridge, MA: MIT Press.

Gupta, A. S., van der Meer, M. A. A., Touretzky, D. S., & Redish, A. D. (2010). Hippocampal replay Is not a simple function of experience. *Neuron*, *65*, 695-705.

Hiratani, N., & Latham, P. E. (2020). Rapid Bayesian learning in the mammalian olfactory system. *Nature Communications*, *11*. doi:ttps://doi.org/10.1038/s41467-020-17490-0

Honig, W. K. (1981). Working memory and the temporal map. In N. E. Spear & R. R. Miller (Eds.), *Information processing in animals: Memory mechanisms* (pp. 167-197). Hillsdale, NJ: Erlbaum.

Hull, C. L. (1930). Knowledge and purpose as habit mechanisms. *Psychological Review*, *37*, 511-525. doi:10.1037/h0072212

Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical Review*, *106*(4), 620-630.

Jaynes, E. T. (2003). *Probability theory: The logic of science*. New York: Cambridge University Press.

Jenkins, H. M., Barnes, R. A., & Barrera, F. J. (1981). Why autoshaping depends on trial spacing. In C. M. Locurto, H. S. Terrace, & J. Gibbon (Eds.), *Autoshaping and conditioning theory* (pp. 255-284). New York: Academic.

Kalmbach, A., Winiger, V., Jeong, N., Asok, A., Gallistel, C. R., Balsam, P. D., & Simpson, E. H. (2021 under review). Mesolimbic dopamine can encode reward availability on multiple time scales without predicting behavior.

Kinney, J. B., & Atwal, G. S. (2014). Equitability, mutual information, and the maximal information coefficient. *Proceedings of the National Academy of Sciences*, *111*(9), 3354-3359. doi:10.1073/pnas.1309933111

Langille, J. J., & Gallistel, C. R. (2020). Locating the engram: Should we look for plastic synapses or information-storing molecules? *Neurobiology of Learning and Memory*, *169.* doi:10.1016/j.nlm.2020.107164

Lattal, K. A., & Gleeson, S. (1990). Response acquisition with delayed reinforcement. *Journal of Experimental Psychology: Animal Behavior Processes*, *16*, 27-39.

Maloney, L. T. (2003). Surface colour perception and environmental constraints. In R. Masufeld & D. Heyer (Eds.), *Colour perception: Mind and the physical world*. New York: Oxford.

Mattar, M. G., & Daw, N. D. (2018). Prioritized memory access explains planning and hippocampal replay. *Nature Neuroscience*, *21*, 1609-1617. doi:10.1038/s41593-018-0232-z

McKenzie, S., & Eichenbaum, H. (2011). Consolidation and reconsolidation: two lives of memories? *Neuron, 71*(2), 224-233. doi:10.1016/j.neuron.2011.06.037

Niv, Y. (2019). Learning task-state representations. *Nature Neuroscience, 22*(10), 1544-1553. doi:10.1038/s41593-019-0470-8

Niv, Y., Daw, N. D., & Dayan, P. (2005). How fast to work: response vigor, motivation and tonic dopamine. In Y. Weiss, B. Schölkopf, & J. R. Platt (Eds.), *NIPS 18* (pp. 1019–1026). Cambridge, MA: MIT Press.

Ólafsdóttir, H. F., Bush, D., & Barry, C. (2018). The Role of Hippocampal Replay in Memory and Planning. *Current Biology, 28*, R37-R50. doi:10.1016/j.cub.2017.10.073

Panoz-Brown, D., Iyer, V., Carey, L. M., Sluka, C. M., Rajic, G., Kestenman, J., . . . Crystal, J. D. (2018). Replay of Episodic Memories in the Rat. *Current Biology, 28*(10), 1628-1634.e1627. doi:https://doi.org/10.1016/j.cub.2018.04.006

Panzeri, S., Harvey, C. D., Piasini, E., Latham, P. E., & Fellin, T. (2016). Cracking the Neural Code for Sensory Perception by Combining Statistics, Intervention, and Behavior. *Neuron, 93*, 491-507.

Poo, M.-m., Pignatelli, M., Ryan, T. J., Tonegawa, S., Bonhoeffer, T., Martin, K. C., . . . Stevens, C. (2016). What is memory? The present state of the engram. *BMC Biology*. doi:10.1186/s12915-016-0261-6

Rescorla, R. A. (1967). Pavlovian conditioning and its proper control procedures. *Psychological Review, 74*, 71-80. doi:10.1037/h0024109

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Systems Technical Journal, 27*, 379-423, 623-656. doi:10.1002/j.1538-7305.1948.tb01338.x

Simoncelli, E. P., & Olshausen, B. (2001). Natural image statistics and neural representations. *Annual Review of Neuroscience, 24*, 1193-1216.

Skinner, B. F. (1938). *The behavior of organisms*. New York: Appleton-Century-Crofts.

Slonim, N., Gurinder, S. A., Tracik, G., & Bialek, W. (2005). Information-based clustering. *Proceedings of the National Academy of Sciences USA, 102*, 18297-18302. Retrieved from **http://www.genomics.princeton.edu/biophysics-theory/Clustering/web-content/index.html**

Stocker, A., & Simoncelli, E. P. (2008). A Bayesian model of conditioned perception. *Advances in Neural Information Processing Systems*, 1490-1501.

Stout, S. C., & Miller, R. R. (2007). Sometimes-competing retrieval (SOCR): a formalization of the comparator hypothesis. *Psychological Review, 114*(3), 759-783. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17638505

Sutton, R. S. (1984). *Temporal credit assignment in reinforcement learning.* (PhD). University of Massachusetts, Amherst, Amherst, MA. Retrieved from http://scholarworks.umass.edu/dissertations/AAI8410337

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement Learning*. Cambrdige, MA: MIT Press Press.

Taylor, K. M., Joseph, V., Zhaoc, A. S., & Balsam, P. D. (2014). Temporal maps in appetitive Pavlovian conditioning. *Behav Processes, 101*, 15-22. doi:doi.org/10.1016/j.beproc.2013.08.015

Thomson, W. l. L. K. (1883). Electrical units of measurement. In *Popular Lectures and Addresses* (Vol. 1, pp. 73-460). New York: Macmillan and Company.

van de Ven, V., Jäckels, M., & De Weerd, P. (2022). Time changes: Timing contexts support event segmentation in associative memory. *Psychonomic Bulletin & Review, 29*(2), 568-580. doi:10.3758/s13423-021-02000-0

Yin, H., Barnet, R. C., & Miller, R. R. (1994). Trial spacing and trial distribution effects in Pavlovian conditioning: Contributions of a comparator mechanism. *Journal of Experimental Psychology: Animal Behavior Processes, 20*, 123-134.

Zentall, T. R. (2019). Rats can replay episodic memories of past odors. *Learning & Behavior, 47*(1), 5-6. doi:10.3758/s13420-018-0340-3