BRILL *Timing & Time Perception* (*2022*) DOI: 10.1163/22134468-bja10069

Timing
&
Time Perception
brill.com/time

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

C. Randy Gallistel[1],* and Peter E. Latham[2]

[1]Professor Emeritus, Rutgers University, 252 7th Ave 10D, New York, NY 10001, USA
[2]Gatsby Computational Neuroscience Unit, Sainsbury Wellcome Centre or Neural Circuits and Behaviour, 25 Howland St., London WIT 4JG, UK
*Corresponding author; e-mail: randy-gallistel@gmail.com
ORCID iD: Gallistel: 0000-0002-4860-5637

Received 31 May 2022; accepted 19 October 2022

Abstract
Bayesian parameter estimation and Shannon's theory of information provide tools for analysing and understanding data from behavioural 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, reinforcement-by-reinforcement and response-by-response basis, important parameters of timing behaviour and of the neurobiological manifestations of timing in the brain. These tools enable quantification of relevant variables in 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) in reinforcement learning. They enable comparing the strength of the evidence for a measurable association to the strength of the behavioural evidence that the association has been perceived. A GitHub site and an OSF site give public access to well-documented Matlab and Python code and to raw data to which these tools have been applied.

Keywords
Bayesian updating, event-by-event parameter estimation, learning rate, Pavlovian conditioning, operant conditioning, reinforcement learning, measuring association, time-scale-invariance

1. Introduction

Information theory and Bayesian approaches to statistics are natural companions. Together, they can assist us in analysing, intuitively understanding, and formally modelling results from experiments that investigate the role of interval timing in

DOI: 10.1163/22134468-bja10069

behaviour and its role in associative learning. It has long been clear that associative learning depends on timing (Gallistel & Gibbon, 2000; Gibbon & Balsam, 1981; Stout & Miller, 2007; Yin et al., 1994). In this paper, we lay out the basics of Bayesian parameter estimation and Shannon's theory of information, as they apply to the behavioural 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 prespecified 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 behaviour and/or in neurobiological activity appear, and how soon evidence of timed responses appears.

The second step in data processing is to make use of those parameters. For that we turn primarily to information theory, for which a fundamental quantity is the entropy. Especially relevant to our analysis is the *entropy difference* (denoted ΔH) between two distributions. We show that this quantity is useful even when the distributions are not accurately known, but are assumed to have the form dictated by the *maximum entropy principle.* This principle is an information-theoretic realization of Occam's razor, assume as little as possible (Jaynes, 1957, 2003).

An example of two distributions used to analyse data from Pavlovian timing experiments are the distribution of inter-reinforcement 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 (a.k.a. operant conditioning) is the distribution of inter-response intervals and the distribution of inter-reinforcement intervals. We show that ΔH 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 (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 a nonlinear change in variable).

A second fundamental quantity in information theory is the *Kullback–Leibler divergence*, denoted by D_{KL}. The D_{KL} is a function that measures the divergence

of one distribution from another of the same form, taking their parameter vectors as inputs.

Our $n\mathrm{D}_{\mathrm{KL}}$ statistic denotes the extent to which a distribution of interest, with sample size n, diverges from a reference distribution. It measures 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 the data 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, $n\mathrm{D}_{\mathrm{KL}}$. In practice, the n derives from the sizes of both the samples from which the rate parameters have been estimated. The $n\mathrm{D}_{\mathrm{KL}}$ is to ΔH as the significance of a correlation coefficient is to the coefficient itself: ΔH measures the statistical association, while the $n\mathrm{D}_{\mathrm{KL}}$ measures the strength of the evidence for it.

The $n\mathrm{D}_{\mathrm{KL}}$ is also a simple, datum-by-datum measure of the strength of the evidence that a parameter of the distribution of a behavioural 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 and respond to the temporal association between a CS and an unconditional stimulus (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 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 $n\mathrm{D}_{\mathrm{KL}}$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 $n\mathrm{D}_{\mathrm{KL}}$ to p value.

Both ΔH and the $n\mathrm{D}_{\mathrm{KL}}$ are computed from estimates of the parameters of the distributions from which the data are assumed to come. In our analyses, these distributions are assumed to be exponential, whether they are or not. This strong simplifying assumption has four justifications:

- It makes ΔH and $n\mathrm{D}_{\mathrm{KL}}$ computable by simple closed-form formulae.
- There is extensive experimental evidence that the learning rate and the difference in performance in associative protocols are primarily determined by the ratio of reinforcement rates (the reciprocals of the mean waits for reinforcement). This dependence implies that the only statistic that matters to the subject in making these decisions is the relative rates of reinforcement. Put another way, the behaviourally 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 they 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.

The approach to associative learning here developed treats association as an objective property of a subject's experience, that is, as a measurable stimulus. These tools measure the strength of the stimulus and the strength of the evidence for it, given the data the subject has seen. The same tools measure the strength of the evidence the subject's behaviour provides as to whether it has perceived the measured association.

2. 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 $n\mathrm{D}_{\mathrm{KL}}$ when plotted as a function of n is a measure of effect size; the greater the divergence between two distributions, the steeper the slope.

Finally, because we are interested in acquisition and extinction and, more generally, in the course of behavioural change, we often want stimulus parameter estimates and behavioural 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 events it experiences in our experiments are manifestations of a stochastic process whose form and parameters the subject must infer from the observable outcomes. The evidence for the form and parameters 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 behavioural 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 behaviour appears? To answer these questions, we measure the strength of the evidence provided by the stimulus and the strength of the evidence provided by the behaviour using the same datum-by-datum statistic.

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 our nonstandard terminology are explained later.

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

3. Distributions

Distributions 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 (Fig. 1). When the support is continuous (in technical language, uncountably infinite), the distribution assigns *probability densities* (Fig. 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

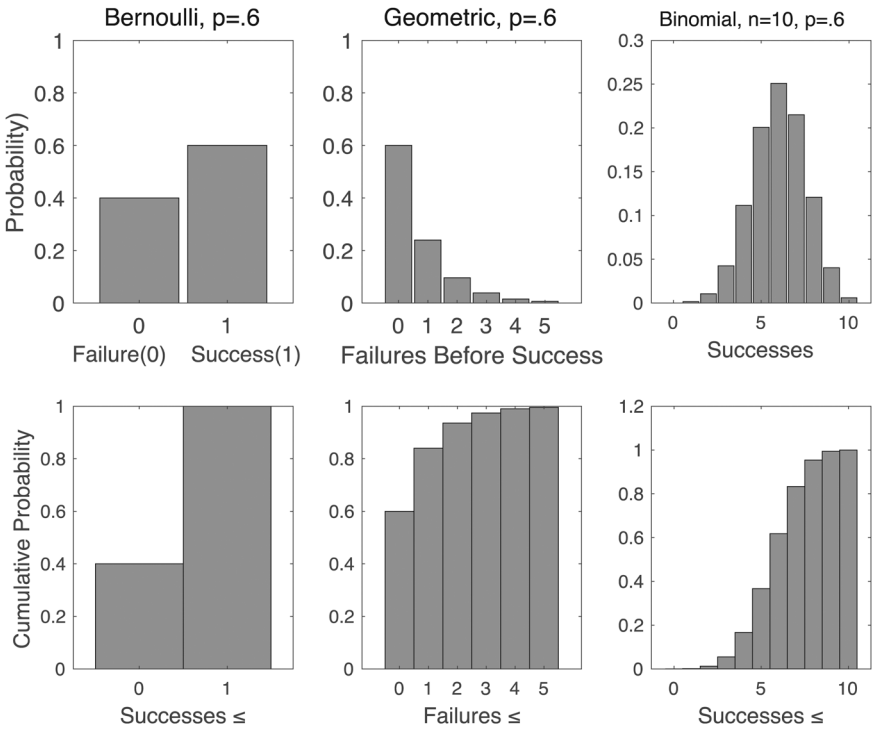

Figure 1. Three common discrete distributions: the Bernoulli, the geometric and the Binomial. They are plotted with bars rather than curves because the support is discrete. Discrete support may always be represented by the integers, as for example in the common practice of representing 'failure' by the 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).

be seen in the second rows of Figs 1 and 2, cumulative distributions asymptote to 1. That is because the support of a probability distribution is a (possibly uncountably infinite) set of mutually exclusive and exhaustive possibilities, so its total mass must be 1. 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 (Fig. 2). Whereas *probabilities* always fall between 0 and 1 (Fig. 1 and Fig. 2 bottom row), *probability densities* (Fig. 2, top row) may take on values from 0 to +infinity. When, for example, the cumulative

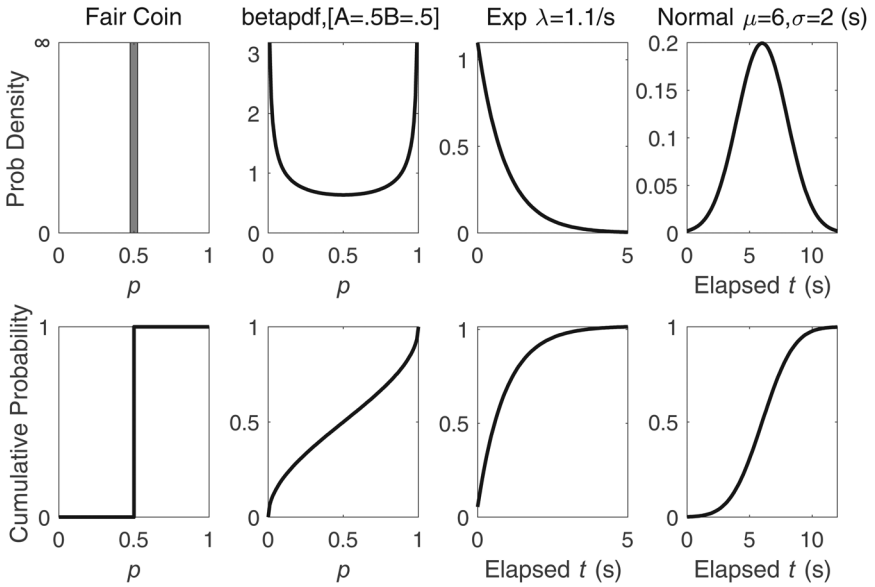

Figure 2. Four common distributions over continuous variables. It is stipulative that a fair coin have a probability of heads of exactly 0.5; therefore, the distribution of the probability of obtaining heads when flipping a fair coins is a vertical line at $p = 0.5$ with no width, infinite height and area (width × 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 = 0.5. 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.

probability function is a step from 0 to 1 at some point along the x-axis (Fig. 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).

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.

3.1. 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 modelling behaviour have only one or two parameters. For example, for Normal distributions, $\theta = [\mu\ \sigma]$, whereas for the Bernoulli, $\theta = p$; and for the Exponential, $\theta = \lambda$, the rate parameter.

3.2. 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 two 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 one-dimensional because these source distributions have only one 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.

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).

3.3. 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 often 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 by including an 'M' (for Model) in the denotation of the prior distribution, but often the 'M' is not included.

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 (Gallistel et al., 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 four 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, the subject 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, as that can substantially reduce the amount of data required to estimate the CoV 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 (its 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 are 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.

3.4. 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, because the form does not change when it is updated. 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' – not one of the distributions that are available in standard scientific programming languages. One has to proceed numerically, which can be tricky and tedious. For example, if one chooses any form for the prior on the Bernoulli other than the beta form, then one has to compute the likelihood function, 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

In summary, using the conjugate form for the prior has several advantages:

– The form of the posterior does not change when new data arrive.
– Therefore, when the prior is updated, only the values of its parameters (the so-called 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 basic sample statistics (usually sums and counts).
– The conjugate prior for a given source distribution, if it exists, is unique.

3.5. 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* (a.k.a. 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. A list of the different parameterizations may be found in the Appendix A1 along with simple custom Matlab™ (MathWorks, Natick, MA, USA) functions implementing the updating functions. These same custom updating functions in both Matlab and Python (Python Software Foundation, Friedricksburg, VA, USA) may be found on at https://github.com/bendecorte/gallistelWorkshop. 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 are very little data (e.g., one 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): $\theta_{beta0} = [0.5\ 0.5]$; $\theta_{gam0} = [0.5\ 0]$; $\theta_{ng0} = [0\ 0\ -0.5\ 0]$. 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

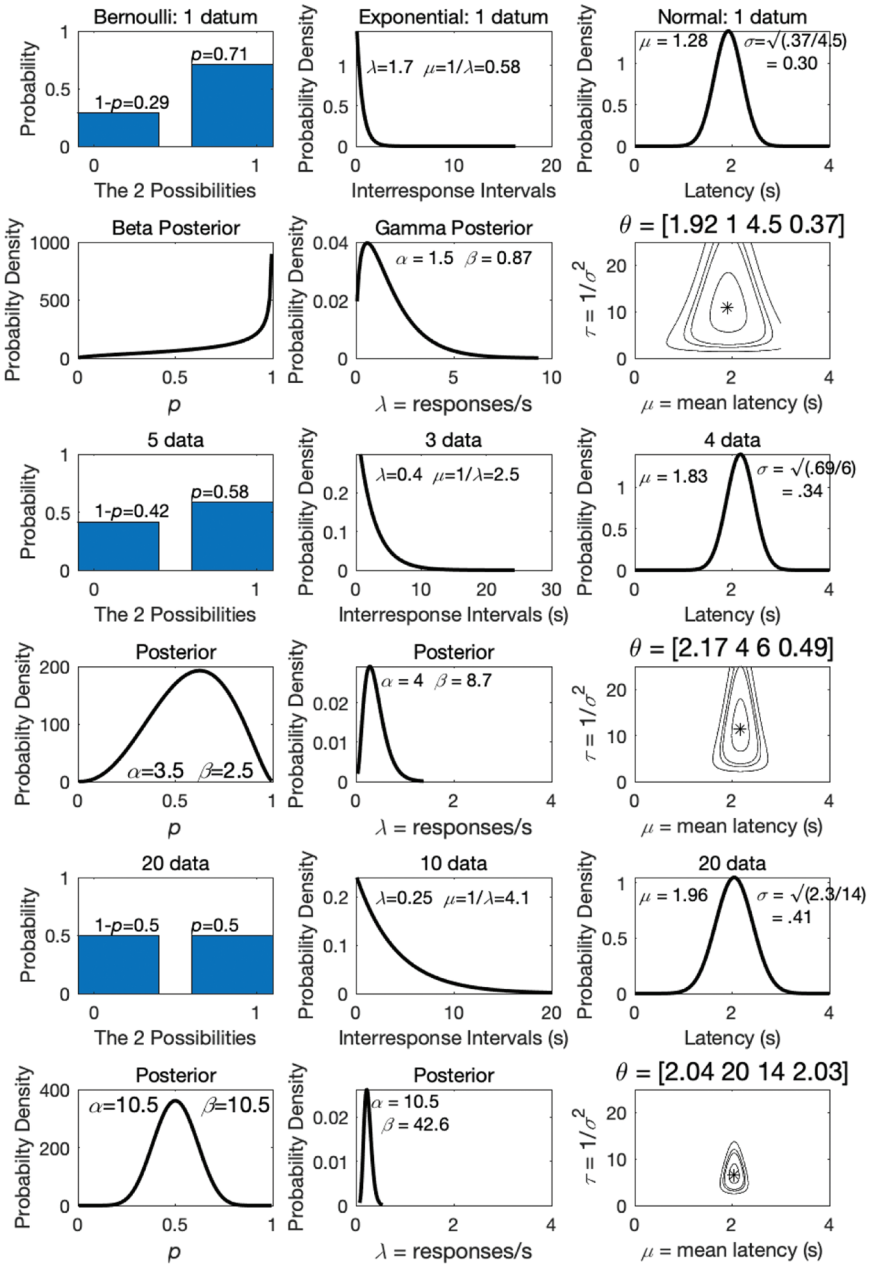

Figure 3. Estimated source distributions (odd rows) and the corresponding conjugate posterior distributions (even rows) for the Bernoulli source (col. 1), the exponential (col. 2), and the Normal (col. 3). The estimate of the source distribution's parameter(s) is shown on each source plot. The updated 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

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 are very little data. However, we, like our subjects, are interested in the conclusions one may rationally draw when there are almost no data.

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 $\theta = [0.5\ 0.5]$, 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 integers 0 and 1 ('failure' and 'success') are the support for the source distribution, 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 Fig. 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 $\theta = [0.5\ 0]$. 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 Fig. 3 plots the estimated Normal distribution and the Normal-gamma posterior distribution on its mean and precision (= 1/var), given

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 in col. 3 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 one 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 one datum and biases later estimates.

1, 4 and 20 draws from a Normal distribution with a mean of 2 and a standard deviation of 0.4 (therefore, a CoV of 0.2).

An *informative prior* was used to illustrate what one might do in estimating a temporal CoV. It was $\theta_0 = [0\ 0\ 4\ 0.36]$. It asserts that, before we have seen any real data, we have seen four 'ghost' data – that 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 fourth element in the informative θ_0 is a suggested sum of squared deviations and the third element is the number of deviations on which this suggestion is notionally based.

For two reasons, the third 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 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 would probably be less than 0.3. Another way of stating that knowledge is that the variance will probably be $0.3^2 \leq 0.09$. (Squaring σ to get the variance is an example of a change-of-variable formula.) To get that variance given an n of 4, the sum of squared deviations has to be $4^*0.09 = 0.36$ (another example of applying a change-of-variable formula).

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.

3.6. Note for Those Who Know Bayes' Rule

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

$$p(\theta|\boldsymbol{D}) \propto p(\boldsymbol{D}|\theta)\pi(\theta) \tag{1}$$

Here, $p(\theta|\boldsymbol{D})$ is the posterior distribution over the parameter vector, θ, given the data, $\boldsymbol{D}$; $p(\theta|\boldsymbol{D})$ is the *likelihood function,* and $\pi(\theta)$ is a prior distribution over the possible values for the parameter vector of the assumed source distribution. While 'likelihood' and 'source distribution' refer to the same function – $p(\boldsymbol{D}|\theta)$ – they mean different things: $p(\boldsymbol{D}|\theta)$ is the source distribution when viewed as a function of the data, $\boldsymbol{D}$, but it is the likelihood when viewed as a function of the parameters, θ.

To take a concrete example, assume that successes and failures are drawn from a Bernoulli distribution with parameter p. Assume also that there have been four draws, and $p = 0.5$. The source distribution is the probability of each of the five possible outcomes – 0, 1, 2, 3 or 4 successes – *given* that $p = 0.5$. These probabilities follow a binomial, and their respective values are 1/16, 4/16, 6/16, 4/16 and 1/16 – which sum to 1, as they should. The likelihood, on the other hand, is the probability of having observed the data – let's say three failures and one success – for all the values p might possibly assume; it is given by $4 \times (1 - p)^3 p$. Unlike source distributions, likelihoods functions do not integrate to 1 (in this example, the integral over p of the likelihood is 1/5), further driving home the point that likelihoods are not probabilities.

It may help to think of $p(\boldsymbol{D}|\theta)$ as a function (command, sub-routine) in a programming language: To compute the likelihood function, you run the function 'backwards' – with the data regarded as parameters and different possible values for θ regarded as the input. In that case, the function delivers as output the *likelihoods* (N.B., not the *probabilities*) of different possible values for the source distribution's parameter. To compute the source distribution, you run the function 'forwards' – with θ regarded as parameters and different values of the data regarded as the input. In both cases the output of the subroutine is commonly denoted by $p(\boldsymbol{D}|\theta)$; what differs is what one thinks of as parameters and what one thinks of as input.

Our expression for Bayes Rule (1) asserts a proportion ($\propto$) not an equality (=). That is because when the right-hand side is integrated over θ (or summed for discrete variables), it does not equal 1, whereas the integral of the left-hand side does (because it is 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 of the product on the right. This product is sometimes called the *marginal likelihood* or the *model evidence*. It is all one needs in computing a point estimate for the source parameter(s) and a credible interval (or contour) on that estimate. That is one reason why the 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.

4. 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 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 is 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 behaviour 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 et al., 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 predictive and retrodictive models of the experienced world.

The preceding two paragraphs presuppose we understand what information *is*. Until, 1948, one could only babble when asked to say what it is. 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.

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

$$H = \sum_{i=1}^{i=n} p_i \log_b \frac{1}{p_i} \tag{2}$$

where p_i is the probability of the ith member of the support set, n is the number of elements 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; they differ only by a scaling factor, $\log_2(\mathrm{e}) \simeq 0.693$. Because entropy depends on $\log(1/p)$, it must be non-negative.

Entropy is a measure of uncertainty: the higher the entropy of a distribution on some empirical variable, the more uncertain one is about the value that variable

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 scarcely, in your thoughts advanced to the stage of science." (Thomson, 1883, p. 72).

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 roll is 1/6 while the chances of correctly guessing the outcome of a coin flip is 1/2 (so you are much more likely to correctly guess the coin than the die). A biased coin has lower entropy than an unbiased one, because you are more likely to correctly guess the outcome (provided you know which call is the better bet). In the extreme case in which the probability of, say heads, is 1, the entropy is zero, and you are sure of the outcome before the coin is flipped.

The entropy of a distribution is the average *surprisal* of a datum drawn from it. The surprisal is $\log(1/p)$, where p is the probability of drawing (observing) that datum. This measure of surprise is intuitive, because the more improbable something is, the more surprised we are when we see it. At the other extreme, because the probability of the sun rising in the morning is 1, our surprise when it happens is 0. Taking the log of the reciprocal of p makes surprisals additive, which is what makes them averageable.

4.1. 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 do not 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 approaches zero, those operations become increasingly accurate. However, one thing we cannot do when we discretize is compute entropy. It is 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, because the possibilities are infinite, and the entropy goes to infinity.

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

$$H[p(x)] = \int dx p(x) \log_b(1/p(x)) \tag{3}$$

This is, typically, finite, but it is no longer the entropy as defined in Equation (1) – the infinite part of the entropy has been thrown away (technically, the part with $\log(dx)$ in it). The problem with throwing away infinities is that it is 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 – a point we shall return to shortly.

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

$$I(x,y) = H[p(x)] - \int dy p(y) H[p(x|y)] \tag{4}$$

where the notation $p(x|y)$ means the probability distribution over x conditioned on y. Information, as defined here, has several features (none of which are especially obvious) that make it especially useful. First, $I(x,y) = I(y,x)$, meaning the information y tells us about x is exactly the same as the information x tells us about y. Second, information is independent of the units with which one measures x and y, and it is even invariant under a nonlinear change of variables. Third, the above definition of mutual information applies to discrete distributions as well (simply replace the integral over y by a sum), and to mixed discrete and continuous distributions

Mutual information is the difference in the entropy of very specific distributions. Is the difference in entropy of an arbitrary pair of distributions also well behaved? It is slightly better behaved than entropy, since it does not depend on units. However, it does change under a nonlinear change of variables. This is of practical importance only when it is not clear how to parameterize quantities of interest, which usually hinges around the question of whether to use log or linear units. But when it is clear which we should use, this is a technical issue we can ignore.

4.1.1. Available Information

Mutual information has a natural interpretation of obvious psychological and neuroscientific importance: it is the average reduction in uncertainty about x one can get from observing y – and vice versa, because $I(x,y) = I(y,x)$. To make this more intuitive, we note that when y is a direct measure of x, but with error bars (i.e., $y = x \pm \Delta x$), then information is approximately equal to the entropy of $p(x)$ with x expressed in the units equal to the error bars, Δx. For instance, if observing y told us the value of x to within 2 cm, then $I(x,y) \approx H[p(x)]$ if x is measured in units of 2 cm. More concretely: if $p(x)$ is uniformly distributed in the range 0–16, and observing y pins down x to within 2 cm, then $H[p(x)]$ is 3 bits $\left[\log_2 8 = \log_2(16/2)\right]$ when x is measured in units of 2 cm. Which makes sense, as there are about eight distinguishable intervals.

Approximating information by entropy measured in units of the error bars 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. It also breaks down if the error bars are not independent of x. However, for small measurement errors – that is, small Δx – compared to the width of $p(x)$, and measurement errors that do not depend on x, it is a good approximation.

For any distribution $p(x)$, we can identify the entropy of $p(x)$, when x is measured in units of our measurement error, as the *available information*. This gives us the intuitively sensible result that the smaller the measurement error, the higher the available information: knowing the value of x to within 1 nanometre will give us much more information than knowing the value of x to within 1 metre. Importantly, the available information gives us a bound on the mutual information, $I(x,y)$ – a bound that is very often not reached. For instance, the GPS on modern phones gives our location to within about 10 metres, whereas by looking at nearby landmarks we know where we are to within about half a metre. Thus, the information provided by the GPS is about 4.3 bits ($\log_2 20$) smaller than the available information. On the flip side, given that we can only locate ourselves to within about half a metre, it would make no sense for GPS to give us our location with smaller error bars – we simply could not make use of that information.

Importantly, the available information is an approximate bound on the mutual information, because a signal cannot communicate more information than is available to be communicated. Put another way, there is no such thing as negative uncertainty.

4.2. 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 to minimize the amount of memory used to store the data coming from some source, the lengths of the code words must be adjusted to make them proportional to the logs of their relative frequencies. 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.

4.3. Measured Divergence

The Kullback–Leibler divergence *of* a distribution, P, *from* a distribution, Q, is denoted $\mathrm{D}_{\mathrm{KL}}(P||Q)$. 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, $\mathrm{D}_{\mathrm{KL}}(P||Q) \neq \mathrm{D}_{\mathrm{KL}}(Q||P)$.

Some information theorists consider the Kullback–Leibler divergence to be a more foundational information-theoretic measure than entropy. Unlike entropy, it is well defined for both continuous and discrete distributions (although it cannot mix the two) and invariant under a change of variables.

When n_Y data have come from a distribution, Y, that diverges from a distribution, X, by $\mathrm{D}_{\mathrm{KL}}(Y||X)$, the cumulative number of memory bits that have been

wasted encoding the *y*s on the assumption they were *x*s is $n\mathrm{D}_{\mathrm{KL}}$. We call this the *cumulative coding cost*.

In our practice, neither Y nor X – the arguments of the D_{KL} – are known exactly. Therefore, both sample sizes must be taken into account. We show in the Appendix A2 that when two distributions with the same form do not differ, then $n = \left(n_y/\left(1 + n_y/n_x\right) \times \mathrm{D}_{\mathrm{KL}}\right.$ is asymptotically distributed $\Gamma\left(n_p/2, 1\right)$, 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). Hereafter, for simplicity, when we say we have computed $n\mathrm{D}_{\mathrm{KL}}$, we did so for $n = n_1/\left(1 + n_1/n_2\right)$. This expression gives the value for the n in $n\mathrm{D}_{\mathrm{KL}}$ that takes proper account of the sample sizes on which the estimates of the probability distributions are based.

The $n\mathrm{D}_{\mathrm{KL}}$ statistic is a simple information-theoretic measure of the strength of stochastic evidence. Unlike a p value, it has physical meaning; it estimates the amount of memory to be saved by recoding the Y data in the light of the evidence that the Y distribution diverges from the X distribution. The uncertainty about the true value of the $n\mathrm{D}_{\mathrm{KL}}$ is conditional on the data – as is an F ratio or p value or any measure of the strength of the evidence. The fewer data, the greater the uncertainty.

5. Measuring Association and Contingency

Events are temporally associated to the extent that the temporal location of the next event may be predicted from knowledge of the location of the preceding event, and vice versa. For events occurring at a given rate, an exponential distribution of event times maximizes entropy (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.

One way to think about the exponential distribution is that in any (infinitesimal) time bin dt, the probability of an event occurring – say the appearance of a food pellet – is $\lambda \times dt$ where λ is the (constant) rate of events. 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 time 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. The distributions of both the forward and backward intervals are exponential with rate λ – exactly the same distribution as the original set of points! (The reason is that we are more likely to drop our pointer onto long intervals.)

Consider a stream of events, $[y_1, y_2, \ldots]$, occurring at times $t_1^y, t_2^y \ldots$., constrained to occur at a fixed rate, λ. As just mentioned, the distribution with the maximum possible uncertainty about t_{n+1}^y given t_n^y is the exponential with rate λ. The more predictable t_{n+1}^y becomes, the more the entropy decreases. When $t_{n+1}^y - t_n^y$ is a constant, t_{n+1}^y is completely predictable when given t_n^y, and the y events are maximally self-associated. In that case the entropy of t_{n+1}^y given t_n^y is minus infinity $[\log(0)]$ – a fact that should not bother us because measurement error, which is always present, will make the entropy finite. In practice, there is never an infinite amount of *available* information.

Consider now a second stream of events, $[x_1, x_2, \ldots]$, occurring at times $t_1^x, t_2^x \ldots$.. 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.

A natural measure of the predictability of the next t^x given a t^y is the conditional entropy, denoted $H(X|Y)$, which is the entropy of $p(t^x|t^y)$ averaged over t^y (the second term in Equation (4), but without the minus sign). High conditional entropy – more uncertainty about the value of t^x given t^y – implies low predictability, and vice versa. Predictability is maximized when t^y predicts t^x to within measurement error, at which point $H(X|Y) = 0$ when t^x is measured in units of the measurement error. Predictability is minimized when t^x tells us nothing about t^y, in which case $H(X|Y) = H(X)$. Similarly, the retrodictability of the preceding t^y given a t^x is maximized when $H(Y|X) = 0$ (again when t^y is measured in units of the measurement error) 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 (within measurement error). The minimization of predictability (maximization of uncertainty) occurs when both distributions are independent. In that case, the mutual information is 0, because $H(X|Y) = H(X) - H(X|Y) = 0 = H(Y) - H(Y|X)$.

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-reinforcement 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 (not CS). That is why we denote the unconditional distribution by C: one can think it means either Chamber or Context.
- In operant conditioning (reinforcement learning), the *prospective* contextual distribution – looking ahead from response to later reinforcement – is

the distribution of the intervals between reinforcements, while the *retrospective* contextual distribution – looking back from reinforcement to the earlier response – is the distribution of inter-response intervals.

– Let Y be the *conditional distribution* of intervals, with rate parameter $\lambda|Y$. In excitatory Pavlovian conditioning, this is the distribution of waits for reinforcement signalled by CS onsets. In inhibitory Pavlovian conditioning and in trace conditioning, it is the distribution of waits for reinforcements signalled by CS offsets. In operant conditioning, the *retrospective conditional distribution* is the distribution of intervals looking back from the reinforcements to the most recent responses. 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 $\lambda|C \le \lambda|Y$, the contextual rate is less than or equal to the conditional rate.
– We treat the contextual and conditional distributions as maximum entropy given the rates, which means we treat them as exponential. Their entropies are thus computed using the formula for the differential entropy of the exponential, $1 - \ln(\lambda)$, where λ is the rate parameter ($\lambda = 1/\mu$).

The proposed measure of association is

$$\Delta H|Y\&C = (1 - \ln(\lambda|C)) - (1 - \ln(\lambda|Y)) = \ln(\ln(\lambda|Y)) - \ln(\ln(\lambda|C)) = \ln\frac{\lambda|Y}{\lambda|C}. \quad (5)$$

Since, as stipulated above, $\lambda|C \le \lambda|Y$, $\Delta H|Y\&C$ is always positive.

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

$$\mathcal{C}(X;Y) = \frac{\Delta H|Y\&C}{I_{\max}} \quad (6)$$

where X denotes the distribution of intervals in the context in which the ys occur (that is, the contextual distribution or the marginal distribution), Y denotes the distribution of the ys, $\Delta H|Y\&C$ is, roughly speaking, the mutual information between the xs and the ys, and $I_{\max}$ denotes the available information – which, as discussed above, depends on measurement error.

In words – using a well-known example – Equation (5) 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 reinforcement following the onset of a CS. Equation (6) says that the contingency is that reduction normalized by the *available information*, the amount of information that a CS could convey. Note that Equation (6) applies more generally than for exponential distributions; in the general case we could replace the numerator with the mutual information and the denominator either with the maximum information or, equivalently, the

entropy of the prior distribution with the relevant variables measured in units of the measurement error.

For rates, we can take $I_{\max}$ to be the entropy of the unconditional entropy measured in units of time corresponding to the interval within which events are perceived to be simultaneous. This, together with the stipulation that the $\lambda_{Y\&C} \leq \lambda_{C,}$ bounds contingency at 1. For human observers, the interval within which events are perceived as simultaneous is measured in tenths of a second (Grabot & van Wassenhove, 2017; Stone et al., 2001; van Wassenhove et al., 2007), with substantial between-subject differences. In the case of anticipatory behaviour, the divisor is the minimum interval within which subjects can make an anticipatory response, such as a peck or a freeze – the interval that is too short to make anticipatory behaviour possible. In rabbit eyeblink conditioning, this interval is 0.1s. In pigeon autoshaping protocols, it is perhaps a second or two.

For a concrete example of a contingency computation, Levinthal et al. (1985) did rabbit eyeblink conditioning with one trial per day and a CS–US interval of 2 s – much longer than the intervals commonly used. Taking 0.1 s as the measurement error, hence the appropriate unit of time, we have I_{max} $I_{max} = \log_2[(60^*60^*24)/0.1)] = 19.7$ bits and $\Delta H|Y\&C = \log_2[(1/2)/1/86400] = 15.4$ bits, so $\mathcal{C}(X;Y) = 15.4/19.7 = 0.78$.

The rabbits learned to blink to the CS in five or six trials. By contrast, when the US–US intervals in rabbit eyeblink conditioning are measured in seconds rather than days, rabbits learn to blink much more slowly, even when the delay of reinforcement is much shorter (see Fig. 10 in Gallistel and Gibbon, 2000). That fact brings us to a discussion of time-scale invariance in association perception.

5.1. The Time-Scale Invariance of Association

The proposed measure of association, Equation (5), 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 reinforcement; it measures a quantitative fact about the temporal distribution of events. The rate ratio in Equation (5) is unitless, and thus time-scale-invariant. There is extensive evidence that Pavlovian conditioning is also time-scale-invariant (Gallistel & Gibbon, 2000), suggesting that it depends on the perception of the association measured by ΔH.

The evidence for time-scale invariance 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 recommended 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 (a.k.a.

operant conditioning). In reinforcement learning, the subject not only learns what predicts and retrodicts what, it also learns a reinforcement-producing or avoiding action. Moreover, its previous behaviour determines whether it has the information necessary to identify a motivationally appropriate action (in RL language, to choose or activate a policy).

In Pavlovian conditioning, the behaviour is irrelevant to the learning process. Its only role is to reveal to the experimenter whether or not the subject has perceived the association. However, behavioural and electrophysiological research has shown that retrospection – looks back in time – occurs even in Pavlovian paradigms (Komura et al., 2001; Matzel et al., 1988; Miller & Barnet, 1993; Namboodiri & Stuber, 2021; Namboodiri et al., 2019; Savastano & Miller, 1998). On the assumption that reinforcing events are rarer than events that *might* (but very often *do not*) predict reinforcements, retrospective interval computation may be the rule rather than the exception. It may be that prospective and retrospective intervals are computed from a temporal map only when something worth predicting actually happens (Arcediano et al., 2003; Balsam & Gallistel, 2009; Chandran & Thorwart, 2021; Honig, 1981). On this hypothesis, all knowledge of temporal intervals derives from looks back in time made possible by a temporal map, a time-stamped record of events. The map makes retrospection possible, just as a spatial map makes navigation possible.

Balsam and Gallistel (2009) suggest that the rate ratio in Equation (5) be called a protocol's *informativeness*, because it determines ΔH, the amount of information a subject may gain from a CS. Gibbon and Balsam (1981) called it the C/T ratio for the following reason: in a pigeon autoshaping protocol, the key on the wall of the chamber is illuminated at more or less random intervals for a fixed duration, at the end of which the food (US) is delivered, regardless of anything the subject does. Each illumination is called a *trial*. Different labs used different trial durations (denoted T) and different US–US intervals (denoted C for cycle duration). The wait for reinforcement after CS onset is T, and the average interval between the termination of the previous trial and the onset of the next is commonly called the intertrial interval or ITI for short; in other words $C = T + ITI$ in pigeon autoshaping.

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 12 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 (Fig. 4): the data are well described by a one-parameter regression equation: $n_R\,(\lambda_R\,|CS/\lambda_R|\,C - 1) = n_R(\text{mean}\,(US - US)\,/\,(\text{mean}\,(CS - US) - 1) = k$, where $k = 294 \pm 28, n_R$ is the number of reinforcements prior to the appearance of a conditioned response and

λ_R is the rate of reinforcement. The learning rate is, by definition, $1/n_R$. The regression model applies over learning rates, from 0 (infinite USs to acquisition) to 1 (acquisition following the first US) – a span of almost three orders of magnitude on both axes. It accounts for 75% of the variance in Fig. 4, with no evidence of systematic deviation, as evidenced by the out-of-sample circles, which were not included in the fit nor in the variance calculation.

The success of the regression model in Fig. 4 is a theory killer. It kills every formal model of associative learning based on delta-rule updating, because it is not reconcilable with the assumption that the updating of associative strength depends on the probabilities of the occurrence and non-occurrence of reinforcements in the presence of a cue. When informativeness is 300, the conditioned response appears after one reinforcement. When informativeness is 1.5, 200 reinforcements do not suffice to make it appear. In both cases, the probability of CS

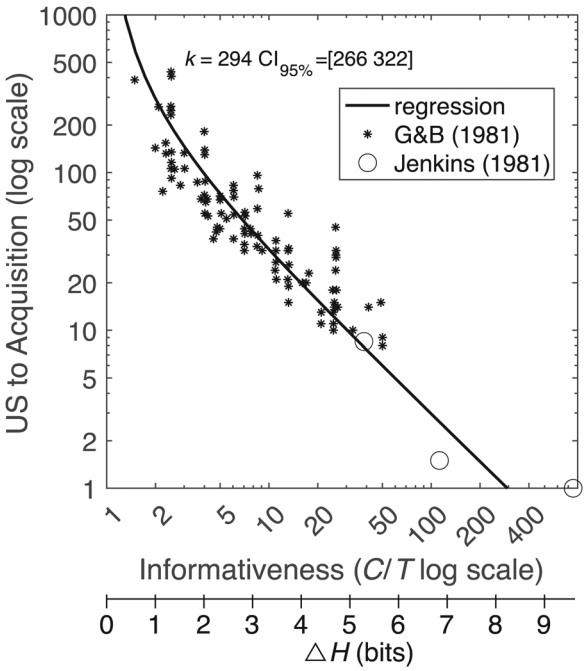

Figure 4. Median unconditional stimuli (US) (reinforcements) to the acquisition of a conditioned response in pigeon autoshaping protocols, as a function of the informativeness (I) of the protocol, on double-logarithmic coordinates. The asterisks are the data plotted in figure 7.11 on p. 245 of Gibbon and Balsam (1981). The regression line was fit to those data. The three open circles are out-of-sample data from Jenkins et al. (1981 Grp 300 Table 8.1 p. 259, Grp W300 Table 8.3 p. 261, and Grp No F Table 8.5 p. 264). The second x-axis shows the relation between informativeness and its logarithm, ΔH, our suggested measure of associative strength. The horizontal deviation from the regression of the open circle on the x-axis was determined by the experimental design, not the subjects.

reinforcement is 1 and the probability of intertrial interval (ITI) reinforcement is 0. We are not aware of a formalized theory of associative learning that can explain the simple quantitative relation between informativeness and the rate of learning. It would appear to be unreconcilable with any 'neurobiologically plausible' model in which synaptic strengths (connection weights) are increased on every reinforced trial and decreased on every unreinforced trial.

The quantity on which the rate of learning depends, $\lambda_R|CS/\lambda_R|C - 1$, where C now stands for context, not cycle duration, is the percent increase in the rate of reinforcement to be expected when the CS is present ($\lambda_R|CS$), relative to the rate expected simply from being in the context in which the CS occurs ($\lambda_R|C$). Thus, when $\lambda_R|CS/\lambda_R|C = 1.5$ there is a 50% increase in reinforcement rate when the CS comes on. That makes intuitive sense on what might be called the make-hay-while-the sun-shines principle, when pecking the key is understood as foraging behaviour (making hay). The parameter k has a data-anchored interpretation; it is the informativeness that produces one-trial learning.

One would like to understand this decision criterion. Figure 5 may give a hint. The nD_{KL} is a measure of the strength of the evidence that $\lambda_R|CS$ differs from $\lambda_R|C$. Figure 5 plots nD_{KL} as a function of the strength of an association, as measured by ΔH, experienced for the first time. When the association has been perceived only once, the effective n in nD_{KL} is 0.5. The dashed vertical is the ΔH that causes one-trial learning in a variety of Pavlovian protocols with a variety of subjects. The equivalent p values are plotted as horizontal dashed lines to give a more intuitively accessible feel for the strength of the evidence at different cumulative coding

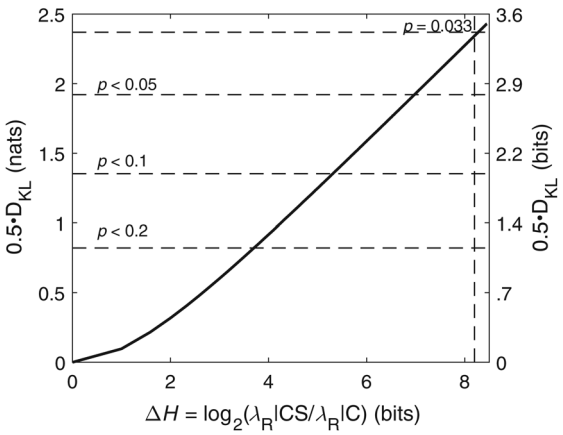

Figure 5. The nD_{KL} following the first experience of a CS-US association (when the sample size is 1 for the conditional and unconditional rate estimates), plotted against the measure of association, $\Delta H|Y\&C = \log_2(\lambda|CS/\lambda|C)$. The vertical dashed line is the association that produces one-trial learning in the median subject. The horizontal dashed lines give the probability of receiving this amount of information from a spurious association.

costs. The amount of evidence the median subject requires to begin responding to a perceived association after a single experience falls comfortably within conventional evidentiary bounds.

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 $0.5 \cdot n_R D_{KL}(\lambda_R|CS, \lambda_R|C) > 2.35 = 0.5 \cdot D_{KL}(k, 1)$, the estimated *cumulative* coding cost when acquisition occurs after a single reinforcement. Thus, on this hypothesis, $n_R|1st\ CR = 4.7/D_{KL}(\lambda_R|CS, \lambda_R|C)$. Near the left end of the regression, when $\lambda_R|CS/\lambda_R|C = 1.5, n_R \simeq 400$; whereas $4.7/D_{KL}(1.5, 1) = 4.7/0.0721 = 65$. The predicted number of reinforcements required for acquisition is off by a factor of 6. Put another way, instead of requiring the 3.5 bits of evidence, which is the required amount for acquisition after a single experience, the median subject requires $1.44 \cdot 400 \cdot 0.5 \cdot D_{KL}(1.5,1)$ bits – more than 20 bits; an implausibly severe evidentiary criterion ($p < 0.000001$).

While the value for the k of the median subject is surprisingly constant for both pigeons and rats, in both excitatory and inhibitory protocols, there is large within-group variation. Some subjects in every group whose data we have seen begin much sooner and some much later than would be expected from the regression equation in Fig. 4. When the mutual information exceeds 4 bits, one subject may begin after a single experience of the association, while another running on the same protocol begins only after 100 experiences. Whatever determines the value of k varies greatly between subjects.

The information communicated by a CS about the wait for a reinforcement cannot be greater than I_{max}, the available information. We see in Fig. 4 that the communicated information that produces one-trial learning in pigeon autoshaping is about 8 bits. In the Jenkins et al. (1981) experiments, the wait for reinforcement after CS onset was 8 s. Therefore, a pigeon's reaction time to a peckable CS is shorter than 8 s. It seems unlikely that it could be shorter than 0.1 s. In principle, the amount of information a subject can get from a cue is limited only by the subject's projected life time and the estimate of its measurement error (the smallest duration that registers). However, when it comes to estimating learning rates, there are analytic limits on the *relevant* amounts of information. The rate should be 0 when the cue provides no information. The results from the experiments using the truly random control suggest that this is so when care is taken to eliminate adventitious associations suggested by small sample noise in early trials (Rescorla, 2000). The empirical equation for the learning rate is also limited by the amount of information that produces one-trial learning and by the interval within which subjects perceive cue and the reinforcement to occur simultaneously.

In addition to the pigeon autoshaping data in Fig. 4 and the rabbit eyeblink data in Gallistel and Gibbon (2000, Fig. 10), we have seen very similar data on inhibitory conditioning in rats. In those protocols, the US occurred at a random rate only during the intertrial intervals, never during the CS. Thus, it was predicted

by CS offset, but the specific time at which it would occur could not be predicted, only its rate of occurrence. Those data come from the laboratory of Peter Balsam and Eleanor Simpson. They are in a paper in the late stages of preparation for publication. Justin Harris has graciously shared his very extensive C/T data from excitatory hopper-entry conditioning in rats. He has presented his results at a scientific meeting and is preparing his data for publication.

Any model of associative learning must confront this startling manifestation of time-scale invariance: the rate of learning depends on a ratio of two average intervals (C and T) – that is, on the informativeness. The duration in the numerator, C, may be three orders of magnitude longer than the duration in the denominator, T. When the informativeness ≈ 1, that is, when $T \approx C$, a conditioned response never appears no matter how often the CS and reinforcement coincide. When C/T is greater than 2, the learning rate is a linearly increasing function of informativeness (C/T) until the learning rate saturates at 1. Saturation occurs when $C/T \approx 300$, in which case a single experience suffices for the median subject to begin responding to the perceived association.

6. Time-Scale Invariance and Contingency in Reinforcement Learning

The dependence of Pavlovian conditioning on the time-scale-invariant $\Delta H|\mathrm{CS}$, which measures the temporal association between cue and reinforcement, poses the question whether the same is true in reinforcement learning (a.k.a. operant conditioning).

Previous work in the information-theoretic framework (Gallistel et al., 2019) implicates the importance of two different associations in reinforcement learning – the prospective association, which is the extent to which responses predict reinforcements, and the retrospective association, which is the extent to which reinforcements retrodict responses. There is a strong *prospective association* when a response communicates substantial information about when to expect reinforcement. There is a strong *retrospective association* when a reinforcement communicates substantial information about the recency of a response.

If the processes that perceive 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 reinforced responses.

Prospective and retrospective associations are not the same, because the distribution of intervals looking back from a reinforcement, R, to the most recent response (r) may be very different from the distribution of intervals looking forward from rs to the next R (Gallistel et al., 2019). (In reinforcement learning – aka operant conditioning – responses are sometimes called “acts” or “operants”

because it is unclear what they are a response to; however, to avoid new notation, we will stick with r for response and R for reinforcement. We will also use r to denote the distribution of intervals to or from a response and R to denote the distribution of intervals to or from a reinforcement.) When pigeons peck a key on conventional variable interval schedules of reinforcement, a peck always precedes reinforcement at a very short fixed interval. Thus, the entropy of the *reinforcement-conditional* distribution of the *retrospective* intervals to a response from a reinforcement, $H\left(r|\overleftarrow{R}\right)$, is 0 (when, as usual, time is measured in units of measurement error). On the other hand, the unconditional (marginal) distribution of inter-response intervals has substantial entropy, because those intervals are approximately exponentially distributed (Gallistel et al., 2019). Therefore, adapting Equation (5) to the present case:

$$\Delta\overleftarrow{H}(r;R) = H(r) - H\left(r|\overleftarrow{R}\right) = H(r) - 0 = H(r) \tag{7}$$

Thus, in this case, a reinforcement communicates all of the available information about the recency of the act that produced it. Therefore, when $H\left(r|\overleftarrow{R}\right) = 0$, the retrospective contingency is 1.

On the other hand, pigeons pecking on variable interval schedules of reinforcement peck at a much higher rate than the rate of reinforcement. Although the delivery of the reinforcement is triggered by a peck, the ineffective pecks between the reinforcements and the reinforcement-triggering peck are so numerous that the distribution of intervals looking forward from pecks to the next reinforcement – the distribution, r, of $r - R$ intervals – is practically indistinguishable from the distribution of $R - R$ intervals (Gallistel et al., 2019). In that case,

$$\Delta\vec{H}(r;R) = H(R) - H(R|\vec{r}) \approx H(R) - H(R) = 0 \tag{8}$$

so the *prospective contingency*, $\Delta\vec{H}(r;R)/H(R)$, approximates 0.

6.1. Degrading the Retrospective Association

Lengthening the hang-fire interval between a reinforcement-triggering act and the reinforcement delivery allows reinforcement-irrelevant acts to intrude into the hang-fire intervals. Their intrusion adds entropy to the retrospective conditional distribution, $r|\overleftarrow{R}$. The longer one makes the hang-fire interval, the greater this entropy becomes; hence, the lower the perceivable retrospective association becomes. Gallistel et al. (2019) found that subjects responding on variable interval (VI) schedules with lengthened hang-fire intervals reduced their rate of response so as to maintain a critical amount of $\Delta\overleftarrow{H}(r;R)$. This result, together with some little-known previous results on instrumental learning with 30-s delays of

reinforcement (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 et al., 2015; Sutton, 1984; Sutton & Barto, 1998). If the credit-assignment process is time-scale-invariant, then the interval between a response and the reinforcement it triggers can be arbitrarily long, provided that the naive response rate is low enough so that the retrospective intervals between initial reinforcements and the responses that trigger them are much shorter than initial estimates of the reinforcement–reinforcement intervals.

A recent experiment in Shahan's lab, now being extended, 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 a 3-s illumination of the feeding hopper signalled 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). A preprint describing the experiment and its analysis, together with the raw data and a spreadsheet with the reinforcement-by-reinforcement estimates of response and reinforcement rates may be found here: https://osf.io/dtnq5/?view_only=9138b977adc344df85657f6bf27aaa41. The estimates of reinforcement rate and response rate are the arguments for the functions that compute the ΔH's – the measures of association – and the $n\mathrm{D}_{\mathrm{KL}}$s – the strength of the evidence for them.

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 3 s 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, the yoked controls 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 four times 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'.

The immediacy supposition is also explicit or implicit 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; Niv, 2019; Niv et al., 2005).

Positing an 'I just made a response' state that endures for 2 minutes after the response seems a stretch. During that delay, rats generate many different responses – and they may make some of them 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 does not 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? Can reinforcement learning be model-free, or must it supervene on a temporal map, the learning of which makes possible the computation of the distributions of prospective and retrospective interval durations?

6.2. Estimating Prospective and Retrospective Associations After the First Few Reinforcements

By Equation (8), the prospective association, $\Delta\vec{H}(r;R)$, is $\log_2\left(\lambda_R|\vec{r}/\lambda_R\right)$, where $\lambda_R|\vec{\mathrm{r}}$ is the response-conditional estimate of the rate of reinforcement [=1/(average wait for reinforcement after making a response)], and λ_R is the marginal (unconditional) rate of reinforcement $[1/(\text{average R} - \text{R interval})]$. Similarly, by Equation (7), the retrospective association, $\Delta\overleftarrow{H}(r;R)$, is $\left(\lambda_r|\overleftarrow{R}\right)/\lambda_r$, where λ_r is the estimated rate of responding and $\lambda_r|\overleftarrow{R}$ is the estimated rate of responding when considering only the intervals looking back from each reinforcement to the most recent response.

The response-conditional rate of reinforcement, $\lambda_R|\vec{\mathrm{r}}$, cannot be less than 0.5/minute given the protocol, because the wait for a reinforcement 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 reinforcements, but they do reduce the average wait between a response and the next reinforcement. Four of the five experimental subjects made additional responses during the 2-minute hang-fire interval after their first response. The closer these additional responses came to the reinforcement triggered by their first response, the shorter the average $r - R$ interval. It was generally less than 2 minutes, particularly early in training. Thus, $\lambda_R|\vec{\mathrm{r}} \geq 0.5/min$ for the experimental subjects. For their yoked controls, on the other hand, the average wait for a reinforcement, after a response, was the average wait from randomly chosen points in time. If the distribution of R – R (reinforcement–reinforcement) intervals is approximately exponential, then the

average wait for a reinforcement from a randomly chosen point in time is equal to the average R – R interval (that is, the contextual inter-reinforcement interval).

How to estimate the marginal distribution (the unconditional distribution of waits for reinforcement) is ambiguous – for us, and probably for the rats as well. They spent 60+ minutes in the test chamber prior to the first reinforced lever press. If one takes the 60 minutes with no reinforcement during context extinction into account, then $\lambda_o < 1/60 \text{ min} = 0.0167 \text{ min}^{-1}$ after the first reinforced lever press. In that case, $\Delta\vec{H}(r;R) = 4.9$ bits.

The ambiguity about the relevant intervals for computing the marginal entropy arises because 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 reinforcement in the first session with the lever present. That latency ranged from 2.8 minutes to 13.6 minutes, yielding unconditional rates of reinforcement of $1/2.8 = 0.36$ to $1/13.6 = 0.074$ reinforcements/minute.

The initial values for the response-conditional rates of reinforcement in this context depend on the initial pattern of responding. For a subject that makes only one response before reinforcement delivery, the initial response-conditional rate of reinforcement is 0.5 min^{-1}. In that case, the prospective ΔH would range from $\log_2(0.5/0.36) = 0.47$ bits to $\log_2(0.5/0.074) = 2.8$ bits. Suppose, however, that a subject makes a first response, waits 110 seconds and then makes nine more responses in the last 10 s prior to reinforcement delivery. The average wait for reinforcement following a response is then $\text{mean}([1\,2\,3\,4\,5\,6\,7\,8\,9\,120]) = 16.5$ s, for a response-conditional rate of reinforcement of $60/16.5 = 3.6 \text{ min}^{-1}$. This rate is much higher than any of the unconditional rates of reinforcement 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 reinforcement triggered by that first response. In that case, the average interval from a response to a reinforcement would be 1 minute, and a response-conditional rate of reinforcement = 1/minute. This dependence of the rates estimates on the number and timing of the interpolated ineffective responses 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 reinforcement do not depend on the subject's behaviour; in operant protocols, they do.

Figure 6a plots the prospective $\Delta\vec{H}(r;R)$ over the first 10 reinforcements for the first pair of yoked subjects. The marginal entropy, $H(r)$, was estimated from reinforcement-by-reinforcement Bayesian estimates of λ_r. The response conditional entropy, $H\left(R|\vec{r}\right)$reinforcement-by-reinforcement estimates of λ_r using only the intervals observed in the context where the lever was present. As always, the

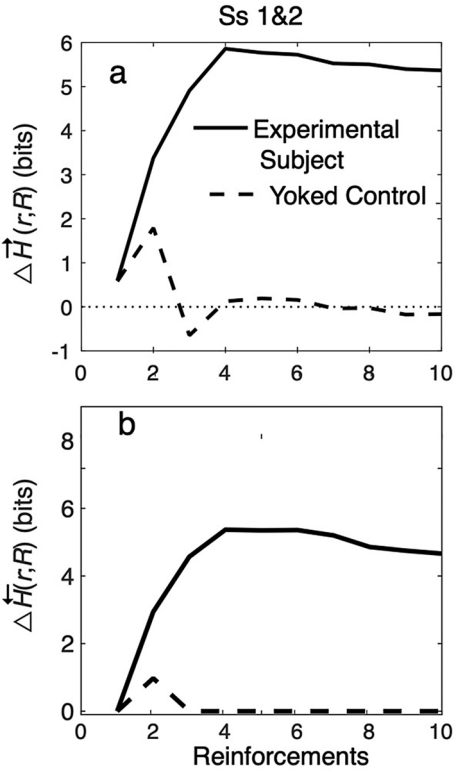

Figure 6. (a) The prospective change in entropy (reduction in uncertainty), $\Delta\vec{H}(r;R)$, as a function of number of reinforcements, for one experimental subject and its yoked control. Only the first 10 reinforcements are shown. (b) The retrospective change in entropy, $\Delta\overleftarrow{H}(r;\mathrm{R})$ for the same pair. The negative ΔHs for the yoked control (dashed curves) are small sample estimation errors; they were asymptotically 0.

unit of time for these rate estimates was chosen so that both were < 1, thereby avoiding negative entropies when using the formula for the differential entropy of the exponential. In this pair of subjects, the ΔH 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 second response; whereas, for the yoked control, it dropped to near 0 after its third response.

Figure 6b plots the retrospective $\Delta\overleftarrow{H}(r;R)$ over the first 10 reinforcements 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 behaviour after a single experience in which there was a 2-minute separation between the causal act and its outcome.

6.3. Estimating the Strength of the Evidence

The main thing a subject wants to know is whether ΔH is positive or zero: if it is positive, then actions affect reinforcements; if it is 0 or negative, actions do not affect reinforcements. The reinforcement-by-reinforcement estimates of the prospective and retrospective associations in Fig. 6 – that is, the ΔHs on the y-axes – were computed by Bayesian estimation of the rate parameters. The estimate after one reinforcement was based on one datum; the estimate after two reinforcements on two data, and so on. But just computing ΔH does not tell a subject whether or not it is statistically different from zero. For that we turn to the $n\mathrm{D}_{\mathrm{KL}}$.

To assess the strength of the evidence for the association they have so far observed subjects need only estimate D_{KL} from their two rate estimates and multiply it by an n derived from the two sample sizes. These computations are much simpler than, for example, the computation of the policy that maximizes the expected cumulative reinforcement (where simplicity is measured by the required number of elementary operations). When the $n\mathrm{D}_{\mathrm{KL}}$ stays near zero, they cannot rule out the possibility that the two rates are the same; when $n\mathrm{D}_{\mathrm{KL}}$ increases linearly, it becomes increasingly likely that the rates are different (see Fig. 5).

Figure 7 plots the $n\mathrm{D}_{\mathrm{KL}}$s for the prospective and retrospective ΔHs against the number of reinforcements. In both cases, the cumulative coding cost for the yoked control is negative at some or even all the plotted points, which seems to contradict that fact that D_{KL} cannot be negative. However, to facilitate graphic interpretation, we have added to the custom functions that compute and plot the $n\mathrm{D}_{\mathrm{KL}}$, an option that allows the user to give the $n\mathrm{D}_{\mathrm{KL}}$ 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 $n\mathrm{D}_{\mathrm{KL}}$s that are in the wrong direction in experimental subjects. Also, the yoked controls' response rate drops to very low values, so the retrospective intervals become very large, which produces conditional entropies greater than the unconditional entropies. In that case, the divergence is in the wrong direction. In looking at $n\mathrm{D}_{\mathrm{KL}}$ graphs, one does not want to confuse these effects with the effects of enduring associations, which always grow steadily greater. Our adding sign to indicate divergences opposite to those expected explains the negative $n\mathrm{D}_{\mathrm{KL}}$s.

7. 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 recent experiment conducted by Basak Akdoğan (Akdoğan et al., in press) in the lab of Peter Balsam and Eleanor Simpson (https://psyarxiv.com/p6v2j).

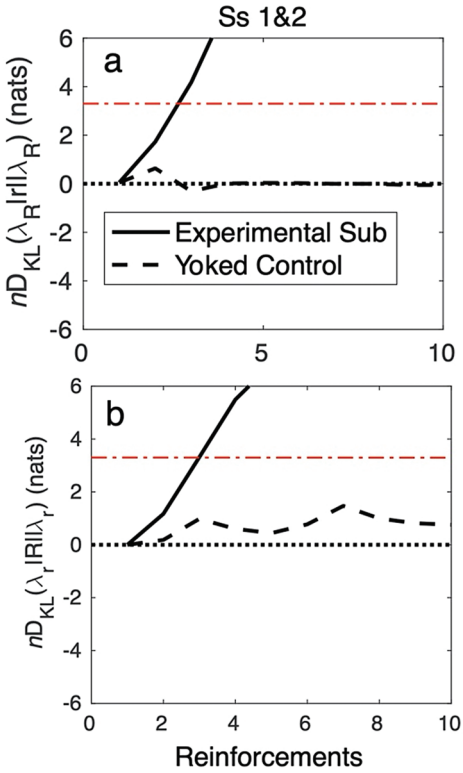

Figure 7. (a) The cumulative coding cost of assuming no prospective association between a response and the wait for reinforcement, as a function of the number of reinforcements, for both the experimental and the yoked subject. When this cost exceeds 3.2 nats (red dashed dot line), the evidence for the association is significant at beyond the 0.01 level. (b) The cumulative coding cost of assuming no retrospective association between a reinforcement and the recency of the last response, as a function of the number of reinforcements, for both the experimental and the yoked subject. For an explanation of how the nD_{KL} acquires its negative sign, see text.

Her experiment used tone durations as the discriminative stimuli (commonly denoted by SΔ in the operant literature), in a two-lever operant choice procedure, with mice as subjects. An SΔ is a signal that indicates which of two possible acts will produce a reinforcement. It is presented just before two levers appear, enabling a choice of actions. In initial training, the SΔ was a tone lasting either 2 s or 6 s. For the subject whose data we analyse here, the choice of the left lever was reinforced following the 2-s tone and the choice of the right lever was reinforced following the 6-s tone.

The subject was pretrained until it chose the correct levers well above chance following both tone durations. When the subject had been responding at asymptote for 600 trials, the SΔs changed: The 2-s tone no longer occurred; it was

replaced by an 18-s tone. The correct response to this novel SΔ was the left lever – the shorter of the two initial SΔs. The 6-s tone continued to occur on 50% of the trials. The correct response on those trials remained what it had always been.

The last 2 s, SΔ occurred on Trial 601; on Trial 602 the SΔ was the already familiar 6 s, the correct response to which was well established. On Trial 603, the SΔ was the novel 18-s tone. The subjects had no way of knowing what the consequences of pressing either lever might now be when given that SΔ. They also had no way of knowing how frequently to expect it. In this new state of the world, there was also no way to know how frequently to expect the other two SΔs (2 s and 6 s) 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 SΔ 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 $\theta_{\text{beta}} = [0.5\ 0.5]$. When updated by the number of correct choices, n_s, and failures to choose correctly, n_f, over the last 300 pre-change 6-s trials, the (hyper)parameters of the beta prior/posterior are $\theta_{\text{beta}} = [n_s + 0.5 \quad n_f + 0.5]$. Figure 8 plots the posterior distribution on the pre-change probability of a correct choice following a 6-s 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 (Bayes & Shannon Code (https://github.com/bendecorte/gallistelWorkshop)). *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 Fig. 8 lies below 0.85 and only 1% lies above 0.93; thus, the odds are 50:1 in favour of the conclusion that the subject's pre-change probability of choosing the right lever was in the interval between [0.85 0.93].

7.1. 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 $n\text{D}_{\text{KL}}$ 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

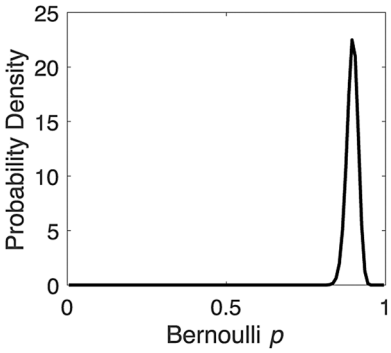

Figure 8. The posterior beta distribution on the estimate of the Bernoulli probability, *p*, of a correct choice given the 6-s discriminative stimulus SΔ before the change in the SΔs.

value and computes the $n\mathrm{D}_{\mathrm{KL}}$ statistic to identify stretches of trials where there is strong evidence of a deviation from the reference distribution:

$$[\mathrm{CmPdif, nDkl, PnDkl, pt}] = \mathrm{BernCCCchange\ (D1, theta1, theta0, 0.9, true)}$$

D1 is the binary vector of successful (reinforced) 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 fifth input) tells the function to plot the figure (see Fig. 9). The fourth input argument, 0.9; is the complement ($1 - \alpha$) of a 'significance' level (α) for the $n\mathrm{D}_{\mathrm{KL}}$ statistic. Including it among the input arguments causes the function to return NaN (not a number) when the number of data and the reference *p* are together such that a 'significant' $n\mathrm{D}_{\mathrm{KL}}$ is impossible. For example, when the α is 0.05 and there are fewer than five data, an $n\mathrm{D}_{\mathrm{KL}}$ significant at α is impossible, because the probability of getting four heads in the first four flips of a fair coin is 0.0625

Sign was added to the plot of the $n\mathrm{D}_{\mathrm{KL}}$ red curve in Fig. 9 to indicate the direction of the divergence, for the reasons described previously. The subject's estimated probability of pressing the right lever following a tone of 6 s duration was lower than the lower limit of the critical intervals on the terminal estimate during Trials 1–8 (when a significant departure of $\hat{p}$ (the estimate) from a reference value of 0.9 was impossible) and then again from Trials 19 to 49, when a significant departure was entirely possible (black curve in Fig. 9). However, the signed $n\mathrm{D}_{\mathrm{KL}}$ reached moderate significance (dashed red horizontal at bottom of plot) on only a single trial (Trial 31, $p < 0.05$). The fact that this trend did not continue, and $n\mathrm{D}_{\mathrm{KL}}$ 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

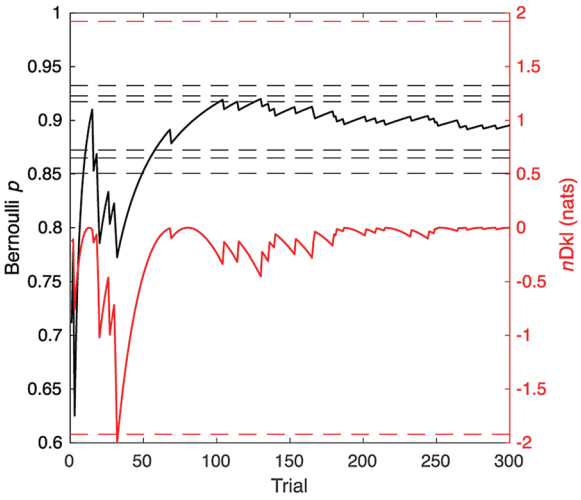

Figure 9. The trial-by-trial estimate of the probability of choosing the right lever as a function of the pre-change sequence of trials (black curve, plotted against left axis). The thin black dashed lines give the critical levels for this estimate, given the complete data set. The red curve is the *signed* $n\mathrm{D}_{\mathrm{KL}}$ statistic, plotted against the right axis. The thin red horizontal dashed lines (bottom and top of panel) represent alpha levels of 0.05 on this statistic.

continues to flip a fair coin, the more certain it becomes that one will observe atypical sequences that seem to imply the coin was not fair over that stretch of flips. Thus, *brief* excursions in $n\mathrm{D}_{\mathrm{KL}}$ beyond essentially arbitrary alpha levels, such as the one that valleys at Trial 31 in Fig. 9, should be ignored. The fluctuations in the $n\mathrm{D}_{\mathrm{KL}}$ already observed between trials 0 and 150 fall well within those expected on the null hypothesis. Therefore, this experiment could have gone on to the next phase much sooner had these analytic methods been used while the experiment was running.

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

7.2. Measuring the Growing Strength of Stochastic Stimuli

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

Similarly, the subject's behaviour is a stochastic stimulus for the experimenter and/or the data analyst: as we observe more behaviour, 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 *behavioural* evidence for contingency perception, as a function of time.

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

As stressed in a previous section, a consideration of fundamental importance in the analysis of *instrumental* behaviour is that the rate at which the subject acquires information about the true state of affairs depends on what reinforcement learning theorists call *exploratory* behaviour, and we call *information-gathering* behaviour. One of the many interesting aspects of Akdoğan'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 Bayesian 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 their behaviour has no effect on the data it has seen (and will see).

The ideal agent, by contrast, tries 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.

There is a vast machine-learning literature on the exploration–exploitation trade-off, with several important mathematical results. However, little research of a quantitative nature has focused on understanding how non-human animals deal with the trade-off between acting so as to gain relevant information and acting so as to maximize return.

A recent machine-learning development of possible relevance to those of us interested in animal behaviour is the emergence of Thompson sampling as one of the leading approaches to the trade-off (Russo et al., 2018). Thompson sampling is otherwise known as *posterior sampling* and as *probability matching.* Probability matching is a striking feature of operant behaviour (Commons et al., 1982; Gallistel et al., 2007; Graf et al., 1964; Herrnstein, 1961; Herrnstein & Loveland, 1975; Maddox & Bohill, 2004). Unlike model-free approaches, Thompson sampling algorithms explicitly model the distributions on the expected loss/gain from each possible action using conjugate priors.

As we have explained, in Bayesian approaches to parameter estimation, posterior distributions are computed as soon as any data are seen and they evolve as more data are seen. These are the tools we need to quantify the evidence our subjects have seen and to quantify the evidence their behaviour provides us as to

the conclusions they have drawn at any given point – the tools required to begin to assess how animal brains deal with the exploration–exploitation trade-off.

7.3. *The Growth of Behaviour-Independent Probability Estimates*

Among the things the subject does not know after the first 18-s tone is the probability of the three different durations so far encountered. Figure 10 plots Bayesian

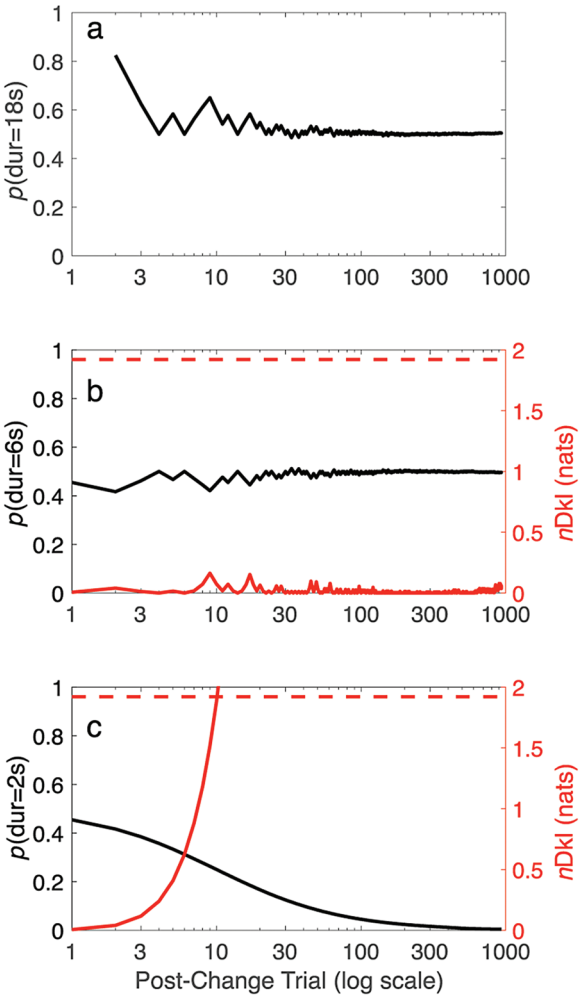

Figure 10. (a) Bayesian estimate of the probability of an 18-s duration tone as a function of the number of trials, counting from its first occurrence. (b) Bayesian estimate of the probability of the 6-s ΔS given a weak presumption that it continues to be 0.5 (black curve, plotted against the left axis) and the (unsigned) $n\mathrm{D}_{\mathrm{KL}}$ 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 0.05 alpha level on the $n\mathrm{D}_{\mathrm{KL}}$. (c) Bayesian estimate of the probability of the 2-s ΔS (black, left axis) and $n\mathrm{D}_{\mathrm{KL}}$ (red, right axis). The odds that it has diminished exceed 20:1 after the 11th post-change trial.

estimates of these probabilities as a function of the number of post-change trials. The estimate for the 18-s tone duration rapidly stabilizes to near its true value (top panel).

The middle panel plots the ideal observer's estimate of the probability of a tone lasting 6s, on the assumption that the first occurrence of the 18-s 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 6-s 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 6-s tone (when estimated using a weakly informative prior) is the same as the old one. The $n\mathrm{D}_{\mathrm{KL}}$ is stable and low, giving no suggestion that this probability has changed.

By contrast, the black curve in the bottom panel of Fig. 10 plots the Bayesian observer's estimate of the probability of a 2-s 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 2-s duration has not occurred. The increasing odds against the no-change hypothesis give reason to abandon the informative prior. When 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 fifth 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.

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

7.4. Tracking the Change in the Behavioural Probabilities

In behaviourist 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 6-s SΔ was reinforced on every 6-s trial both before and after the substitution of an 18-s SΔ for the 2-s SΔ. Its reaction to this change shows that choosing the right lever following a 6-s tone was not a habit; it depended on the arithmetic relation between the three SΔs. (Our focus in what follows is on the tools, not this conclusion; for the full force of Basak's data, see the preprint https://psyarxiv.com/p6v2j.)

In reaction to the appearance of 18-s SΔs and the disappearance of 2-s SΔs, the mouse reduced its probability of pressing the right lever following the 6-s SΔ, even though that response to that stimulus continued to be unfailingly reinforced (Fig. 11a). The reduced probability of pressing the right lever and the correspondingly increased probability of pressing the left lever following 6-s tones became evident on the 14th post-change trial, which was the seventh 6-s trial following the first occurrence of an 18-s tone. The behavioural change is indicated by the downward inflection in the black curve and the corresponding sharp upward inflection in the red curve in Fig. 11a. The odds confirming the existence of this

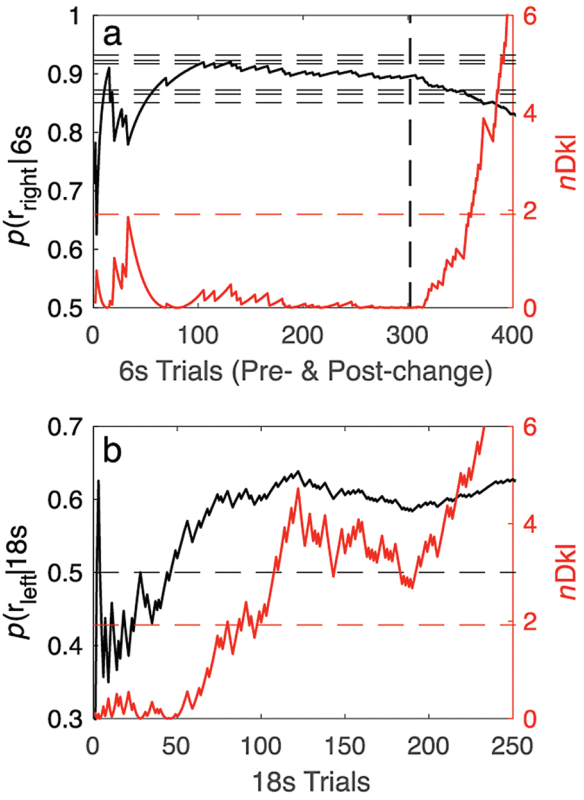

Figure 11. (a) The conditional probability of sampling (pressing) the right lever as a function of all the trials (pre- and post-change) on which the ΔS was 6 s (black curve plotted against left axis) and the cumulative cost of assuming the post-change probability equals the pre-change (in nats, red 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 0.8, 0.9 and 0.98 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 18-s tone. (b) Same plots for the novel 18-s tone.

change permanently exceeded 20:1 after the next 15 6-s trials (red curve, Fig. 11a). At that point, the Bayesian estimate of the probability of choosing the right lever had dropped below the lower 0.01 boundary of the 98% critical interval on the Bayesian estimate of the pre-change probability of this choice (black curve, Fig. 11a).

As may be seen from Fig. 11b, the subject began to respond correctly to the novel 18-s tone long before it recovered its 'habitual' response to the 6-s tone. The enduring disruption of the correct response to the 6-s stimulus reduced the subject's overall rate of reinforcement. It was striking in seven of the eight mice in this protocol. Two of them reduced their probability of pressing the correct lever to below chance and kept it there for hundreds of trials.

One mouse, in strong contrast, did not reduce at all its high probability of pressing the correct lever following the 6-s tone, strongly implying that it *construed* the same experience differently from the other mice. It construed the 6-s tone as the tone lasting 6 s.

In 'model-free' models of brain-mediated reinforcement learning, little attention has been paid to the effects of the different ways in which the same experience (the same state) may be *construed* and the impact of that variable on the effects of changing circumstances (non-stationarity). Under some *construals*, no change in policy may be necessary, whereas under others, it may be. We define a *construal* to be a choice among different representations of its past experience available to a subject. Subjects that represent durations by arithmetically manipulable *numerons* (neurobiological numerals, see Gallistel, 2018) may construe their experience in different ways – both before experiencing a change and/or retrospectively, in the light of information gained after the change.

Many other mice, in both this condition and other conditions, construed the 6-s tone as the shorter of two tones. That construal facilitated the discovery of the optimal action in some of Basak's conditions and hindered it in others (https://psyarxiv.com/p6v2j).

The sustained and substantial reduction in the post-change probability of choosing the right lever following a 6-s 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 behaviour. When the 18-s ΔS supplants the 2-s ΔS, for all the mouse knows, pressing the left lever following a 6-s ΔS may sometimes yield a bigger reinforcement than that yielded by pressing the right lever. Continuing to press the left lever only very rarely on 6-s trials will retard the forming of an estimate of what those two probabilities might be – the probable size of a possibly bigger reinforcement and the probability of producing it. Thus, the rationality/optimality of a subject's post-change behaviour 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 6-s

trials relative to the value it places on maintaining the previously experienced rate of reinforcement on those trials.

8. Measuring Contingency Detection Behaviourally and Photometrically

Kalmbach et al. (2022) measured mesolimbic dopamine activity photometrically in mice that had previously learned to press a lever for food reinforcement. The photometric monitoring of dopamine activity began when these mice first began to hear tones that lasted 80 s, during which lever presses did not produce reinforcements. 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 intervals, the CS and the $\sim$ CS intervals. The $\sim$ CS intervals are usually called the ITIs. When calculating associative strength, the conditional distribution must always be the distribution whose rate of reinforcement is higher than the contextual rate of reinforcement. Thus, the conditional distribution in this protocol is the distribution of US–US intervals during the ITIs. Its rate parameter is $\lambda\,(\mathrm{US}|\sim \mathrm{CS})$, the informativeness is $\lambda\,(\mathrm{US}|\sim \mathrm{CS})\,/\lambda\,(\mathrm{US}|\mathrm{C})$.

Figure 12 plots the trial-by-trial response rate estimates and the $n\mathrm{D}_{\mathrm{KL}}$ for two subjects. In Fig. 12a, the subject began to respond at a higher rate during the ITIs than during the tones only after 300 trials. In Fig. 12b, the subject consistently responded at a higher rate during the ITIs after the eighth trial. This 37-fold difference in the rate of learning is an example of the variability commonly seen in this statistic (trials to acquisition).

To delimit the training interval within which the conditioned behaviour 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 behaviour (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 behaviour 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 the 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 Fig. 12.

Figure 13 plots the signed cumulative coding cost of assuming that the CS rates are the same as the ITI rates for the eight subjects in the negative-contingency ('inhibitory') protocol (top two rows) and the four subjects in the truly random control (no contingency). For the four subjects in the non-contingent condition

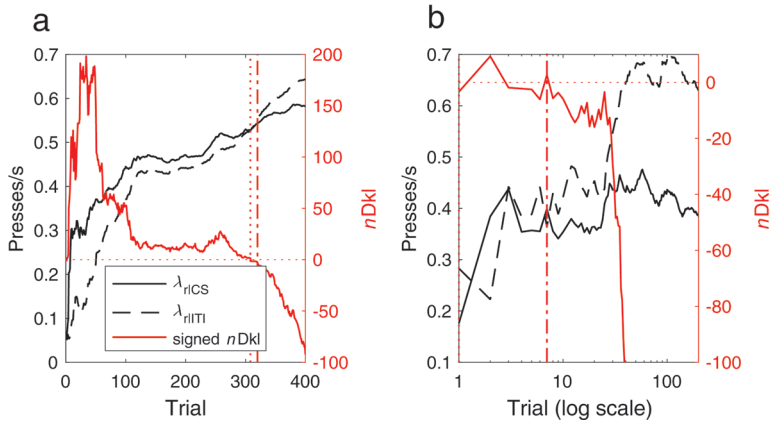

Figure 12. Trial-by-trial estimates of the response rates during the conditional stimuli (CSs) and the intertrial intervals (ITIs) (black solid and dashed curves, left axis) and the signed cumulative cost of assuming that the CS rate is the same as the ITI rate (red curve, right axis). Vertical dotted red lines indicate the trial after which $(\lambda_r|\text{CS} - \lambda_r|\text{ITI})$ is enduringly negative. Vertical dash-dot red lines are the trials after which the *signed* $n\text{D}_{\text{KL}}$ remained less that 3.3 nats (the $p < 0.01$ 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 $n\text{D}_{\text{KL}}$ is permanently less than 3.3 nats ($p < 0.01$). Therefore, the vertical red dash-dot line superposes on the vertical red dotted line. The x-axis has been logged to better reveal what happened over the first 10 trials. (Recall that we assign to the $n\text{D}_{\text{KL}}$ the sign of the difference in response rates.)

(bottom row of Fig. 13), the $n\text{D}_{\text{KL}}$ was positive throughout training. Note also that these $n\text{D}_{\text{KL}}$s did not continue to climb, unlike $n\text{D}_{\text{KL}}$s for the negative-contingency subjects, which maintained or often increased their downward slope as training continued. The slope of the $n\text{D}_{\text{KL}}$ is proportionate to the difference in the rate estimates. When the slope is 0, so is the difference in the rate estimates.

8.1. Applying the $n\text{D}_{KL}$ to the Photometric Data on Dopaminergic Activity

Abby Kalmbach recorded dopaminergic 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. In these subjects, the onset of the CS signalled a decrease in the rate of reinforcement to below the contextual rate and its offset signalled an increase to above the contextual 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.

To measure trial-by-trial the development of the photometric spikes, we constructed templates for them by averaging the same 1.5-s initial and final segments across the last 200 training trials, when the spikes were well developed. We then

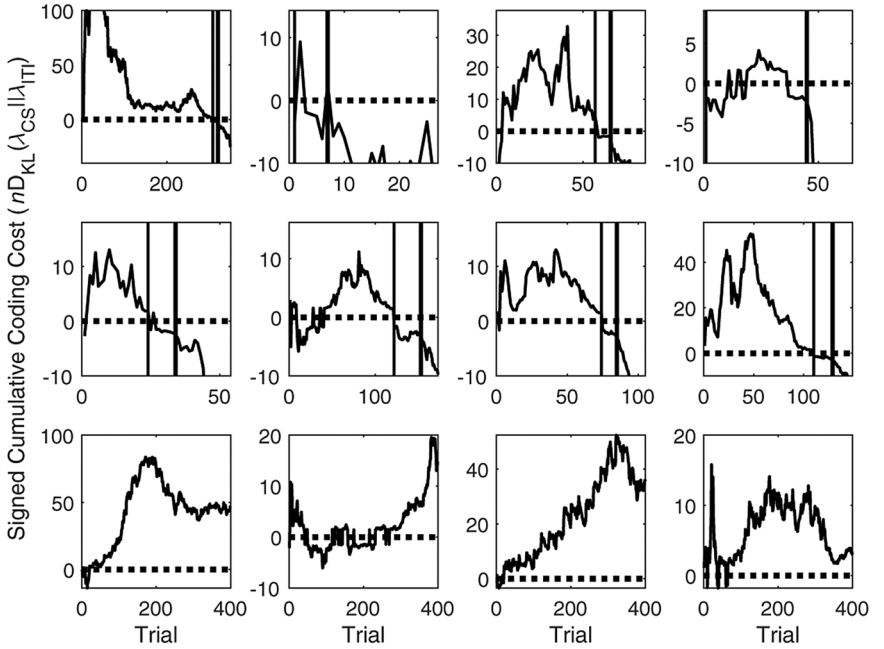

Figure 13. The signed cumulative coding cost (in nats) of assuming that the rate of pressing during the conditional stimuli (CSs) is the same as the rate during the ITIs for all 12 subjects. The thinner vertical line indicates the trial at which the response appeared; the thick vertical line, the trial at which the evidence for it became decisive.

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 – 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 four correlations (third element of the parameter vector) and that the sum of their squared deviations from the mean correlation was 0.9 (fourth element). This prior biased the variance

estimate toward what we knew must be about the right value, thereby heightening the sensitivity of the $n\mathrm{D}_{\mathrm{KL}}$. The $n\mathrm{D}_{\mathrm{KL}}$ 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.

Figure 14 plots the photometric $n\mathrm{D}_{\mathrm{KL}}$s (right two columns) alongside the (negatively signed) behavioural 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 behavioural evidence for the detection of the negative contingency between the CS and reinforcement delivery. However, in one subject, decisive behavioural evidence appeared very quickly and well before decisive photometric evidence (see row 2 in Fig. 14). In all the subjects, the behavioural evidence rapidly got very much stronger than the photometric evidence, because the behavioural '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 *n*Dkls in the right two columns. The strengthening of the photometric signals was, however, less pronounced than the strengthening in the behavioural signals.

9. Conclusions

A temporal map of past experience enables the replay of episodes and the recovery of associative structure (Gupta et al., 2010; Mattar & Daw, 2018; Ólafsdóttir et al., 2018; Panoz-Brown et al., 2018; van de Ven et al., 2022; Zentall, 2019). Information-theoretic tools quantify associative structure by ΔH, which is the information conveyed by the CS in Pavlovian protocols. In reinforcement learning protocols, it can be the prospective information about the expected wait for reinforcement conveyed by a response and/or the retrospective information about the recency of a response conveyed by a reinforcement. The rates, which are the inverses of the mean waits, are computed on the maximum entropy assumption, which is that the distributions are exponential.

Bayesian parameter estimation enables us to estimate ΔH, which measures the strength of these temporal associations, after the first US in the Pavlovian protocol, and after the first reinforced response in operant protocols. The $n\mathrm{D}_{\mathrm{KL}}$ (cumulative coding cost) measures the strength of the evidence for the association. The ΔH is the information-theoretic analogue of a correlation coefficient (but with a range from 0 to +infinity), while $n\mathrm{D}_{\mathrm{KL}}$ is the information-theoretic analogue of its statistical significance, but with a range from 0 to +infinity.

The $n\mathrm{D}_{\mathrm{KL}}$ 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

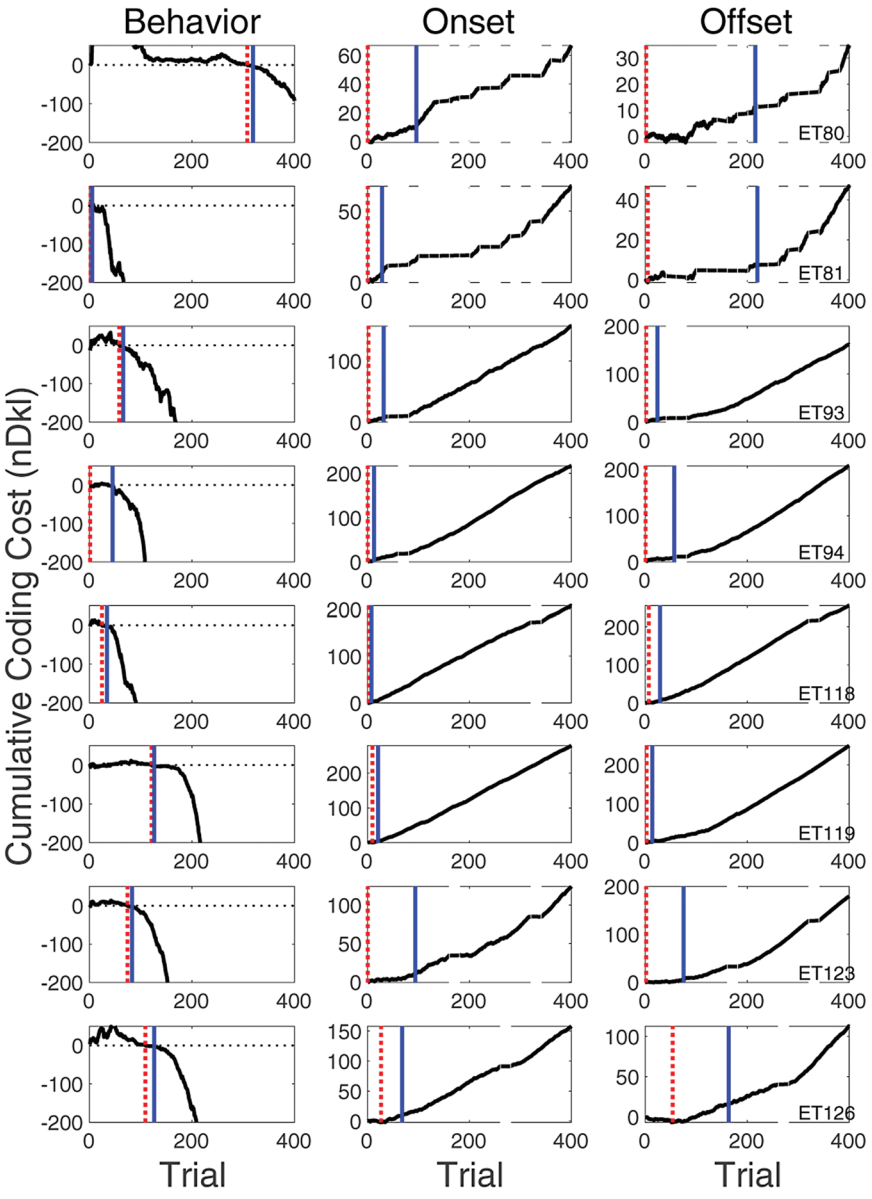

Figure 14. The signed $n\mathrm{D}_{\mathrm{KL}}$ plots (in nats) for the behavioral 'signal' alongside the (unsigned) photometric $n\mathrm{D}_{\mathrm{KL}}$ss for the onset and offset spikes in the dopaminergic photometry signal. Blue verticals mark the trials where evidence becomes decisive; dotted 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 $n\mathrm{D}_{\mathrm{KL}}$s were similar to the behavioral ones (bottom row of Figure 13), in that there was little evidence for the spikes at conditional-stimulus (CS) onset and offset in the photometric signals from the control subjects.

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). If memory capacity is an important resource, brains should recode the data when they discover the associative structure in it.

Adopting improved stochastic models to conserve memory resources also improves a brain's ability to anticipate future reinforcements and punishments and to recognize the causal effects of the behaviour 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 behaviour in Pavlovian conditioning and to operant behaviour 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 reinforcement-by-reinforcement basis the strength of the evidence that a subject's ongoing 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 the associative structure of the experimental environment, these tools put the study of timed behaviour and associative learning on the same conceptual footings as the study of sensory processing and perception. In those fields Bayesian inference and information theory play fundamental roles (Brainard, 2009; Chater et al., 2006; Feldman, 2016, 2021; Froyen et al., 2015; Ganguli & Simoncelli, 2016; Hiratani & Latham, 2020; Maloney, 2003; Panzeri et al., 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 behavioural and neurobiological changes induced by the perception of the association; and (iv) the strength of the evidence for the behavioural and neurobiological 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.

Given a contextual distribution of expected waits, X, and a distribution, Y, of waits conditioned on an event that occurs within that context, the contingency

between Y and X is the information that the y events convey about the expected waits, $\Delta H|Y$, normalized by the available information, $I_{\max}$:

$$\mathcal{C}(X;Y) = \frac{\Delta H|Y}{I_{\max}} \tag{9}$$

where both the numerator and the denominator are computed from the reciprocals of the mean waits using the formula for the differential entropy of an exponential distribution: $H = 1 - \ln(\lambda)$.

The numerator in Equation (9) is well defined because it is the log of the unitless ratio of a conditional rate and the contextual (unconditional) rate, always chosen such that the conditional rate is greater than or equal to the unconditional rate. The temporal unit used in computing the entropies must be the interval within which a subject judges two events to have occurred simultaneously. The choice of a temporal unit has no effect on the numerator in Equation (9), because $\Delta H|Y$ is the log of a unitless rate ratio; it does, however, strongly affect $I_{\max}$, because it is a differential entropy.

In this approach to associative learning, an association is not an associative bond in the subject's mind or brain – not a connection weight nor a Hebbian synapse. Nor is it a subjective value placed on reinforcement. 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 the events. The temporal map enables a brain to look back in time to compute the intervals and the rate parameters of distributions that it implicitly treats as exponential.

Our use of the entropy difference as a measure of temporal association is related to a more general approach to defining clusters of events information-theoretically (Slonim et al., 2005). Events are temporally associated when they cluster in time (or in the frequency domain, which is 1/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). It also provides evidence for some causal process that explains the cluster. Clustering is a time-scale-invariant phenomenon, because mutual information has units determined by the base of the logarithm, not by the units that attach to distributional parameters.

Acknowledgements

We are grateful for data supplied to us for use in this paper by Basak Ardogan in the laboratory of Peter Balsam and Eleanor Simpson at Columbia University and the New York State Psychiatric Institute and by Timothy Shahan at the Utah State University. Research in the Balsam and Simpson lab is supported by RO1MH068073. Research in Tim Shahan's lab is supported by R01HD093734. We are grateful to Ben De Corte for creating the GitHub website where the publicly

accessible Matlab™ and Python code is to be found, for improvements to the Matlab code and its documentation and for the creation of the Python code. We are grateful to Basak Ardogan for extremely helpful feedback on early drafts.

References

Akdoğan, B., Wanar, A., Gersten, B. K., Gallistel, C. R., & Balsam, P. D. (in press). Temporal encoding: Relative and absolute representations of time guide behavior. *J. Exp. Psychol. Anim. Learn. Cogn.* doi: 10.31234 /0sf.io/p6v2j.

Arcediano, F., Escobar, M., & Miller, R. R. (2003). Temporal integration and temporal backward associations in human and nonhuman subjects. *Learn. Behav., 31*, 242–256. doi: 10.3758/BF03195986.

Balsam, P. D., & Gallistel, C. R. (2009). Temporal maps and informativeness in associative learning. *Trends Neurosci., 32*, 73–78. doi: 10.1016/j.tins.2008.10.004.

Brainard, D. H. (2009). Bayesian approaches to color vision. In M. S. Gazzaniga, E. Bizzi, L. M. Chalupa, S. T. Grafton, T. F. Heatherton, C. Koch, J. E. LeDoux, S. J. Luck, G. R. Mangan, J. A. Movshon, H. Neville, E. A. Phelps, P. Rakic, D. L. Schacter, M. Sur, & B. A. Wandell (Eds), *The cognitive neurosciences* (pp. 395–408). Cambridge, MA, USA: MIT Press.

Chandran, M., & Thorwart, A. (2021). Time in associative learning: a review on temporal maps. *Front. Hum. Neurosci., 15*, 617943. doi: 10.3389/fnhum.2021.617943.

Chater, N., Tenenbaum, J. B., & Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. *Trends Cogn. Sci., 10*, 287–291. doi: 10.1016/j.tics.2006.05.007.

Commons, M. L., Hernnstein, R. J., & Rachlin, H. (1982). *Quantitative analyses of behavior. Vol.* 2: *Matching and maximizing accounts*. Cambridge, MA., USA: Ballinger.

Dayan, P., & Niv, Y. (2008). Reinforcement learning: The Good, The Bad and The Ugly. *Curr. Opin. Neurobiol., 18*, 185–196. doi: 10.1016/j.conb.2008.08.003.

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

Feldman, J. (2021). Mutual information and categorical perception. *Psychological Science, 32*, 1298–1310. doi: 10.1177/0956797621996663.

Ferster, C. B., & Skinner, B. F. (1957). *Schedules of reinforcement*. East Norwalk, CT, USA: Appleton–Century–Crofts. doi: 10.1037/10627-000.

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

Gallistel, C. R. (2018). Finding numbers in the brain. *Philos. Trans. R. Soc. Lond. B Biol. Sci., 373*, 20170119. doi: 10.1098/rstb.2017.0119.

Gallistel, C. R. (2021). Robert Rescorla: time, information and contingency. *Rev. Hist. Psicol., 42*, 7–21.

Gallistel, C. R., & Gibbon, J. (2000). Time, rate, and conditioning. *Psychol. Rev.*, *107*, 289–344. doi: 10.1037/0033–295X.107.2.289.

Gallistel, C. R., King, A., & McDonald, R. (2004). Sources of variability and systematic error in mouse timing behavior. *J. Exp. Psychol. Anim. Behav.Process., 30*, 3–16. doi: 10.1037/0097-7403.30.1.3.

Gallistel, C. R., King, A. P., Gottlieb, D., Balci, F., Papachristos, E. B., Szalecki, M., & Carbone, K. S. (2007). Is matching innate? *J. Exp. Anal. Behav., 87*, 161–199. doi: 10.1901/jeab.2007.92-05.

Gallistel, C. R., Craig, A. R., & Shahan, T. A. (2019). Contingency contiguity, and causality in conditioning: Applying information theory and Weber's Law to the assignment of credit problem. *Psychol. Rev., 126*, 761–773. doi: 10.1037/rev0000163.

Ganguli, D., & Simoncelli, E. P. (2016). Neural and perceptual signatures of efficient sensory coding. arXiv:1603.00058. doi: 10.48550/arXiv:1603.00058.

Gershman, S. J., Norman, K. A., & Niv, Y. (2015). Discovering latent causes in reinforcement learning. *Curr. Opin. Behav. Sci., 5*, 43–50. doi: 10.1016/j.cobeha.2015.07.007.

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, NY, USA: Academic Press.

Grabot, L., & van Wassenhove, V. (2017). Time order as psychological bias. *Psychological Science, 28*, 670–678. doi: 10.1177/0956797616689369.

Graf, V., Bullock, D. H., & Bitterman, M. E. (1964). Further experiments on probability-matching in the pigeon. *J. Exp. Anal. Behav., 7*, 151–157. doi: 10.1901/jeab.1964.7-151.

Grünwald, P. D. (2007). *The minimum description length principle*. Cambridge, MA, USA: 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. doi: 10.1016/j.neuron.2010.01.034.

Herrnstein, R. J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. *J. Exp. Anal. Behav., 4*, 267–272. doi: 10.1901/jeab.1961.4-267.

Herrnstein, R. J., & Loveland, D. H. (1975). Maximizing and matching on concurrent ratio schedules. *J. Exp. Anal. Behav., 24*, 107–116. doi: 10.1901/jeab.1975.24-107.

Hiratani, N., & Latham, P. E. (2020). Rapid Bayesian learning in the mammalian olfactory system. *Nat. Commun., 11*, 3845. doi: 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, USA: Erlbaum.

Hull, C. L. (1930). Knowledge and purpose as habit mechanisms. *Psychol. Rev., 37*, 511–525. doi: 10.1037/h0072212.

Jaynes, E. T. (1957). Information theory and statistical mechanics. *Phys. Rev., 106*, 620–630. doi: 10.1103/PhysRev.106.620.

Jaynes, E. T. (2003). *Probability theory: the logic of science*. New York, NY, USA: 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, NY, USA: Academic Press.

Kalmbach, A., Winiger, V., Jeong, N., Asok, A., Gallistel, C. R., Balsam, P. D., & Simpson, E. H. (2022). Dopamine encodes real–time reward availability and transitions between reward availability states on different timescales. *Nat. Commun., 13*, 3805. doi: 10.1038/s41467-022-31377-2.

Kinney, J. B., & Atwal, G. S. (2014). Equitability, mutual information, and the maximal information coefficient. *Proc. Natl Acad. Sci. U. S. A., 111*, 3354–3359. doi: 10.1073/pnas.1309933111.

Komura, Y., Tamura, R., Uwano, T., Nishijo, H., Kaga, K., & Ono, T. (2001). Retrospective and prospective coding for predicted reward in the sensory thalamus. *Nature, 412*, 546–549. doi: 10.1038/35087595.

Langille, J. J., & Gallistel, C. R. (2020). Locating the engram: Should we look for plastic synapses or information–storing molecules? *Neurobiol. Learn. Mem., 169*, 107164. doi: 10.1016/j.nlm.2020.107164.

Lattal, K. A., & Gleeson, S. (1990). Response acquisition with delayed reinforcement. *J. Exp. Psychol. Anim. Behav. Process., 16*, 27–39. doi: 10.1037/0097-7403.16.1.27.

Levinthal, C. F., Tartell, R. H., Margolin, C. M., & Fishman, H. (1985). The CS–US interval (ISI) function in rabbit nictitating membrane response conditioning with very long intertrial intervals. *Anim. Learn. Behav., 13*, 228–232. doi: 10.3758/BF03200014.

Maddox, W. T., & Bohill, C. J. (2004). Probability matching, accuracy maximization, and a test for the optimal classifier's independence assumption in perceptual categorization. *Percept. Psychophys., 66*, 104–118.

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

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

Matzel, L. D., Held, F. P., & Miller, R. R. (1988). Information and expression of simultaneous and backward associations: implications for contiguity theory. *Learn. Motiv., 19*, 317–344. doi: 10.1016/0023-9690(88)90044–6.

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

Miller, R. R., & Barnet, R. C. (1993). The role of time in elementary associations. *Curr. Dir. Psychol. Sci., 2*, 106–111. doi: 10.1111/1467–8721.ep10772577.

Namboodiri, V. M. K., & Stuber, G. D. (2021). The learning of prospective and retrospective cognitive maps within neural circuits,. *Neuron, 109*, 3552–3575. doi: 10.1016/j.neuron.2021.09.034.

Namboodiri, V. M. K., Otis, J. M., van Heeswijk, K., Voets, E. S., Alghorazi, R. A., Rodriguez–Romaguera, J., Mihalas, S. & Stuber, G. D. (2019). Single–cell activity tracking reveals that orbitofrontal neurons acquire and maintain a long-term memory to guide behavioral adaptation. *Nat. Neurosci.*, *22*, 1110–1121. doi: 10.1038/s41593-019-0408-1.

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

Niv, Y., Daw, N., & Dayan, P. (2005). How fast to work: Response vigor, motivation and tonic dopamine. In Y. Weiss, B. Schölkopf, & J. Platt (Eds), *Advances in Neural Information Processing Systems 18* (*NIPS 2005*) (pp. 1019–1026). Cambridge, MA, USA: MIT Press.

Ólafsdóttir, H. F., Bush, D., & Barry, C. (2018). The role of hippocampal replay in memory and planning. *Curr. Biol., 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., Gentry, M., Brotheridge, S., Somekh, I., Corbin, H. E., Tucker, K. G., Almeida, B., Hex, S. B., Garcia, K. D., Hohmann, A. G., & Crystal, J. D. (2018). Replay of episodic memories in the rat. *Curr. Biol., 28*, 1628–1634. doi: 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. doi: 10.1016/j.neuron.2016.12.036.

Poo, M.–m., Pignatelli, M., Ryan, T. J., Tonegawa, S., Bonhoeffer, T., Martin, K. C., Rudenko, A., Tsai, L.-H., Tsien, R. W., Fishell, G., Mullins, C., Gonçalves, J. T., Shtrahman, M., Johnston, S. T., Gage, F. H., Dan, Y., Long, J., Buzsáki, G., & Stevens, C. (2016). What is memory? The present state of the engram. *BMC Biol.*, *14*, 40. doi: 10.1186/s12915-016-0261-6.

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

Rescorla, R. A. (2000). Associative changes with a random CS-US relationship. *Q. J. Exp. Psychol. Compar. Physiol. Psychol., 53B*, 325–340. doi: 10.1080/713932736.

Russo, D. J., Roy, B. V., Kazerouni, A., Osband, I., & Wen, Z. (2018). A tutorial on Thompson sampling. *Found. Trends Mach. Learn., 11*, 1–96. doi: 10.1561/2200000070.

Savastano, H. I., & Miller, R. R. (1998). Time as content in Pavlovian conditioning. *Behav. Process., 44*, 147–162. doi: 10.1016/S0376-6357(98)00046-1.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Techn. J., 27*, 379–423, 623–656. doi: 10.1002/j.1538–7305.1948.tb01338.x, 10.1002/j.1538–7305.1948.tb00917.x.

Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representations. *Annu. Rev, Neurosci., 24*, 1193–1216. doi: 10.1146/annurev.neuro.24.1.1193.

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

Slonim, N., Atwal, G. S., Tračik, G., & Bialek, W. (2005). Information–based clustering. *Proc. Natl Acad. Sci. U. S. A., 102*, 18297–18302. doi: 10.1073/pnas.0507432102.

Stocker, A. A., & Simoncelli, E. (2008). A Bayesian model of conditioned perception. In J. Platt, D. Koller, Y. Singer, & S. Roweis (Eds), *Advances in Neural Information Processing Systems 20* (*NIPS 2007*) (pp. 1490–1501).

Stone, J. V., Hunkin, N. M., Porrill, J., Wood, R., Keeler, V., Beanland, M., Port5, M., & Porter, N. R. (2001). When is now? Perception of simultaneity. *Proc. Biol. Sci., 268*, 31–38. doi: 10.1098/rspb.2000.1326.

Stout, S. C., & Miller, R. R. (2007). Sometimes-competing retrieval (SOCR): a formalization of the comparator hypothesis. *Psychol. Rev., 114*, 759–783. doi: 10.1037/0033-295X.114.3.759.

Sutton, R. S. (1984). *Temporal credit assignment in reinforcement learning.* PhD thesis. Amherst, MA, USA: University of Massachusetts.

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

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

Thomson, W. (1883). Electrical units of measurement. In *Popular Lectures and Addresses* (Vol. *1*, pp. 73–460). New York, NY, USA: Macmillan and Company.

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

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia, 45*, 598–607. doi: 10.1016/j.neuropsychologia.2006.01.001.

Yin, H., Barnet, R. C., & Miller, R. R. (1994). Trial spacing and trial distribution effects in Pavlovian conditioning: Contributions of a comparator mechanism. *J. Exp. Psychol. Anim. Behav.Process.*, *20*, 123–134. doi: 10.1037/0097-7403.20.2.123.

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

Appendix A1

Mathematical Forms for the Conjugate Priors and Their Hyperparameters

Mathematical forms of the posteriors:

(1) The *beta distribution* for the Bernoulli source distribution, when it is parameterized by $\theta_B = p$
(2) The *gamma distribution* for the exponential source distribution, with $\theta_E = \lambda$
(3) The *normal-gamma* distribution for the Gaussian, with $\theta_G = [\mu\, \tau]$, where $\tau =$ precision $=$ 1/variance

The hyperparameter vectors, that is, the parameters of these conjugate prior distributions, are:

(1) $\theta_{beta} = [A\ B]$, where A and B are both *shape* parameters
(2) $\theta_{gam} = [\alpha\, \beta]$, where α is the *shape* parameter and β is the *inverse scale* or *rate* parameter. (In Matlab™, $\theta_E = [A\ B]$, where A is the shape parameter and B the *scale* parameter. This necessitates transforming rate to its inverse.)
(3) $\theta_{ng} = [\mu\, \nu\, \alpha\, \beta]$ aka [mu nu alpha beta], where ν (nu) is a pseudo-n, α is a different pseudo-n and β is a pseudo-sum of squared deviations (further explanation below)

Change of variable formulae (with common parameter names and denotations)

Bernoulli

$O = p/(1 - p)$, where O is the odds ratio, e.g., 2:1
$p = O/(O + 1)$

Exponential

$\lambda = 1/\mu$(rate)
$\mu = 1/\lambda$(mean)

Normal ('mean' a.k.a. 'average' a.k.a. 'location' is almost always denoted by μ)

$\sigma = \sqrt{\text{var}} = 1/\sqrt{\tau}$ (standard deviation)
$\text{var} = \sigma^2 = 1/\tau$ (variance)
$\tau = 1/\text{var} = 1/\sigma^2$ (precision)

Gamma

shape (commonly denoted by k or α or A)
scale (commonly denoted by θ or β or B) $= 1/\lambda$
$\lambda =$ *rate or inverse scale*) $= 1/$*scale*

Normal-gamma

μ ('location')

λ or ν ('lambda' or often 'nu', which is Greek small n. The latter denotation indicates that the initial value assigned to this hyperparameter is a 'ghost' n. Like the 0.5s that occur in the Jeffreys priors for the Bernoulli and the exponential; the initial value for this parameter adjusts the n used to compute the mean of the source distribution)

α the initial value assigned to this hyperparameter is also a 'ghost' n; it alters the value used in computing the variance from the sum of squared deviations

β the initial value assigned to this hyperparameter is a ghost sum of squared deviations.

Update Formulae

(1) $\theta_{\mathrm{Bu}} = [n_{\mathrm{s}} + A_0 \quad n_{\mathrm{f}} + B_0] = [n_{\mathrm{s}} n_{\mathrm{f}}] + \theta_0$, where n_{s} and n_{f} are the numbers of failures and successes in the data. For the custom Matlab™ function named *BayesEstBernP.m* see Bayes&ShannonCode. [The custom Matlab and Python functions here mentioned have been created by CRG (Matlab) and by Ben DeCorte (Python); they cannot be found on the Matlab user website.]

(2) $\theta_{\mathrm{Eu}} = [n_{\mathrm{s}} + \alpha_0 \quad T + \beta_0] = [n_{\mathrm{s}}\ T] + \theta_0$, where n_{s} is the sample n and T is the cumulative duration of observation (the sum of the vector of inter-event intervals whose first element is 0). The Matlab™ custom update function, *expoupdate.m,* is in Bayes&ShannonCode.

(3) $\theta_{\mathrm{Gu}} = [\mu_{\mathrm{u}}\, \nu_{\mathrm{u}}\, \alpha_{\mathrm{u}}\, \beta_{\mathrm{u}}]$,

$\mu_{\mathrm{u}} = (\nu_0 \cdot \mu_0 + n \cdot \bar{x}) / (\nu_0 + \mu_0)$

$\nu_{\mathrm{u}} = n + \nu_0$

$\alpha_{\mathrm{u}} = n/2 + \alpha_0$

$\beta u = \beta 0 + 0.5\left\{n \cdot \sigma + \left[\nu_0 \cdot n \cdot (\bar{x} - \mathrm{mu})^2\right] / (\nu_0 + n)\right\}$,

where (xbar = sample mean; σ = sample sigma; n = sample n)

The custom Matlab™ update function is *normalgamma_update.m*

Appendix A

Distribution of the Kullback-Leibler divergence under maximum likelihood

The problem

Suppose we have two data sets, $\mathrm{x}_i^{(1)}, i = 1, \ldots, n_1$ and $\mathrm{x}_i^{(2)}, i = 1, \ldots, n_2$. We'll assume they're both generated from the same family, which is parameterized by θ. We want to know if they came from the same distribution; that is, if they were generated from the same value of θ. Our approach is to compute the maximum

likelihood estimate of θ for both of them, write down a test statistic, and compute its distribution under the null hypotheses that both data sets came from the same distribution.

Distribution of the maximum likelihood estimate

We'll start by computing the distribution of the maximum likelihood estimate of θ. For that we can drop the superscript telling us which data set it is, but we'll restore it when we generate the test statistic. For that we follow the usual steps. Using D to denote the data,

$$D = \{x_i, x_2, \ldots, x_n\}. \tag{A.1}$$

the log likelihood, $L(\theta)$, is given by

$$L(\theta) = \log P(D|\theta). \tag{A.2}$$

The maximum likelihood estimate of θ, denoted, θ_{ML} is found from by

$$\frac{\partial L(\theta)}{\partial \theta}\Big|_{\theta=\theta_{ML}} = 0. \tag{A.3}$$

Letting $\theta = \theta^* + (\theta_{ML} - \theta^*)$, where θ^* is the value of θ that generated the data, we have

$$\frac{\partial L(\theta^*)}{\partial \theta^*} + (\theta_{ML} - \theta^*) \cdot \frac{\partial^2 L(\theta^*)}{\partial \theta^* \partial \theta^*} \approx 0 \tag{A.4}$$

where "$\cdot$" is the usual dot product. As usual, we'll assume that the second derivative self averages,

$$\frac{\partial^2 L(\theta^*)}{\partial \theta^* \partial \theta^*} \approx \left\langle \frac{\partial^2 L(\theta^*)}{\partial \theta^* \partial \theta^*} \right\rangle \tag{A.5}$$

where the angle brackets indicate an average over the data conditioned on θ^*; that is, an average with respect to $P(D|\theta^*)$. We thus have

$$\theta_{ML} - \theta^* \approx -\left\langle \frac{\partial^2 L(\theta^*)}{\partial \theta^* \partial \theta^*} \right\rangle^{-1} \cdot \frac{\partial L(\theta^*)}{\partial \theta^*}. \tag{A.6}$$

We'll assume that θ_{ML} is Gaussian, so all we need is its mean and covariance. The right hand side should be zero on average, so the mean value of θ_{ML} is θ^*. Its variance is given by

$$\langle(\theta_{\mathrm{ML}} - \theta^*)(\theta_{\mathrm{ML}} - \theta^*)\rangle \approx \left\langle \frac{\partial^2 L(\theta^*)}{\partial\theta^*\partial\theta^*} \right\rangle^{-1} \cdot \left\langle \frac{\partial L(\theta^*)}{\partial\theta^*} \frac{\partial L(\theta^*)}{\partial\theta^*} \right\rangle \cdot \left\langle \frac{\partial^2 L(\theta^*)}{\partial\theta^*\partial\theta^*} \right\rangle^{-1}. \tag{A.7}$$

To cast this expression into a simpler, and more familiar, form we assume, as usual, that the data is independent. In that case,

$$P(D|\theta) = \prod_{i=1}^{n} P(\mathrm{x}_i|\,\theta) \tag{A.8}$$

and so the log likelihood may be written

$$L(\theta) = \sum_{i=1}^{n} \log P(\mathrm{x}_i|\,\theta). \tag{A.9}$$

Defining

$$\mathrm{J} \equiv -\int d\mathrm{x}\, P(\mathrm{x}|\,\theta^*)\, \frac{\partial^2 \log P(\mathrm{x}|\,\theta^*)}{\partial\theta^*\partial\theta^*} \tag{A.10}$$

(note the minus sign, which is needed to make J positive definite), it's easy to see that

$$\left\langle \frac{\partial^2 L(\theta^*)}{\partial\theta^*\partial\theta^*} \right\rangle = -n\mathrm{J}, \tag{A.11}$$

and not much harder to see (via integration by parts) that

$$\left\langle \frac{\partial L(\theta^*)}{\partial\theta^*} \frac{\partial L(\theta^*)}{\partial\theta^*} \right\rangle = n\mathrm{J}. \tag{A.12}$$

Consequently, Eq. (A.7) becomes

$$\langle(\theta_{\mathrm{ML}} - \theta^*)(\theta_{\mathrm{ML}} - \theta^*)\rangle \approx \frac{1}{n}\mathrm{J}^{-1}. \tag{A.13}$$

Alternatively,

$$P(\theta_{\mathrm{ML}}|\,\theta^*) \approx \frac{1}{\mathrm{Det}(2\pi(n\mathrm{J})^{-1})^{1/2}} e^{-n(\theta_{\mathrm{ML}}-\theta^*)\cdot \mathrm{J}\cdot(\theta_{\mathrm{ML}}-\theta^*)/2} \tag{A.14}$$

where Det denotes determinant.

The test statistic

For the test statistic, which we denote z, we'll use the Kullback-Leibler divergence between the two distributions parameterized by their maximum likelihood estimates,

$$z \equiv m \int dx\, P\left(\mathrm{x}\left|\theta_{\mathrm{ML}}^{(1)}\right.\right) \log \frac{P\left(\mathrm{x}|\,\theta_{\mathrm{ML}}^{(1)}\right)}{P\left(\mathrm{x}|\,\theta_{\mathrm{ML}}^{(2)}\right)} \tag{A.15}$$

where m is, for now, a placeholder; it will be chosen later to make the distribution of z as simple as possible. Taylor expanding around θ^* and working to second order in $\theta_{\mathrm{ML}}^{(1)} - \theta^*$ and $\theta_{\mathrm{ML}}^{(2)} - \theta^*$, we arrive at the somewhat surprising result

$$z \approx \frac{m}{2}\left(\theta_{\mathrm{ML}}^{(1)} - \theta_{\mathrm{ML}}^{(2)}\right)\cdot \mathrm{J} \cdot \left(\theta_{\mathrm{ML}}^{(1)} - \theta_{\mathrm{ML}}^{(2)}\right). \tag{A.16}$$

Our goal now is to find the distribution of z. For that we use the usual expression,

$$\begin{aligned} P(z) \approx \int & \frac{d\theta_{\mathrm{ML}}^{(1)}\, d\theta_{\mathrm{ML}}^{(2)}}{\mathrm{Det}\left(2\pi(n_1^{1/2} n_2^{1/2}\mathrm{J})^{-1}\right)} e^{-n_1\left(\theta_{\mathrm{ML}}^{(1)}-\theta^*\right)\cdot \mathrm{J}\cdot\left(\theta_{\mathrm{ML}}^{(1)}-\theta^*\right)\Big/ 2 - n_2\left(\theta_{\mathrm{ML}}^{(2)}-\theta^*\right)\cdot \mathrm{J}\cdot\left(\theta_{\mathrm{ML}}^{(2)}-\theta^*\right)\Big/ 2} \\ & \times \delta\left(z - m(\theta_{\mathrm{ML}}^{(1)} - \theta_{\mathrm{ML}}^{(2)})\cdot \mathrm{J}\cdot (\theta_{\mathrm{ML}}^{(1)} - \theta_{\mathrm{ML}}^{(2)})\Big/ 2\right) \end{aligned} \tag{A.17}$$

where $\delta(.)$ is the Dirac delta distribution. To simplify this expression, we'll let

$$\theta_{\mathrm{ML}}^{(1)} = \theta^* + \mathrm{J}^{-1/2} \cdot \mathrm{y}^{(1)} \tag{A.18a}$$

$$\theta_{\mathrm{ML}}^{(2)} = \theta^* + \mathrm{J}^{-1/2} \cdot \mathrm{y}^{(2)}. \tag{A.18b}$$

Under this change of variables, Eq. (A.17) becomes

$$P(z) \approx \int \frac{d\mathrm{y}^{(1)} d\mathrm{y}^{(2)}}{\left(2\pi \mathrm{n}_1^{1/2}\, \mathrm{n}_2^{1/2}\right)^d} e^{-n_1 \mathrm{y}^{(1)}\cdot \mathrm{y}^{(1)}/2 - n_2 \mathrm{y}^{(2)}\cdot \mathrm{y}^{(2)}/2}\, \delta\left(z - m(\mathrm{y}^{(1)} - \mathrm{y}^{(2)}) \cdot (\mathrm{y}^{(1)} - \mathrm{y}^{(2)})/2\right) \tag{A.19}$$

where d is the dimension of θ; that is, the number of parameters needed to describe $\mathrm{P}(\mathrm{x}|\,\theta)$. Although it won't obviously help at this point, we compute the moment generating function of $\mathrm{P}(\mathrm{z})$, denoted $\tilde{\mathrm{P}}(\mathrm{s})$,

$$\tilde{P}(s) = \int_0^\infty dz\, e^{-sz} P(z) \approx \int \frac{d\mathrm{y}^{(1)} d\mathrm{y}^{(2)}}{\left(2\pi \mathrm{n}_1^{1/2}\mathrm{n}_2^{1/2}\right)^d} e^{-n_1 \mathrm{y}^{(1)}\cdot \mathrm{y}^{(1)}/2 - n_2 \mathrm{y}^{(2)}\cdot \mathrm{y}^{(2)}/2 - ms\left(\mathrm{y}^{(1)} - \mathrm{y}^{(2)}\right)\cdot\left(\mathrm{y}^{(1)} - \mathrm{y}^{(2)}\right)/2}. \tag{A.20}$$

The integrals over $\mathrm{y}^{(1)}$ and $\mathrm{y}^{(2)}$ are reasonably straightforward, and we arrive at

$$\tilde{P}(s) \approx \frac{1}{\left(1 + s\left(\frac{m}{n_1} + \frac{m}{n_2}\right)\right)^{d/2}}. \tag{A.21}$$

This is the moment generating function of the Gamma distribution,

$$\int_0^\infty dy\, e^{-sy} \frac{\beta^\alpha y^{\alpha-1} e^{-\beta y}}{\Gamma(\alpha)} = \frac{1}{(1 + s/\beta)^\alpha}. \tag{A22}$$

Using this expression, we see that, at least in the limit of large n_1 and n_2 (where the Gaussian approximation is good),

$$z \sim \Gamma(d/2, 1/(m/n_1,\ m/n_2)); \tag{A.23}$$

that is, z is approximately Gamma distributed with shape parameters $d/2$ and scale parameter $m/n_1 + m/n_2$.

A convenient choice for m is the one that makes the scale parameter equal to 1, which is

$$m = \frac{n_1}{1 + n_1/n_2}. \tag{A.24}$$

Thus, if we set (in a slight abuse of notation)

$$z \equiv \frac{n_1}{1 + n_1/n_2} \cdot \int dx\, P\left(x \left| \theta_{ML}^{(1)} \right.\right) \log \frac{P\left(x \left| \theta_{ML}^{(1)} \right.\right)}{P\left(x \left| \theta_{ML}^{(2)} \right.\right)}, \tag{A.25}$$

then z will be approximately Gamma distributed with shape parameter $d/2$ and scale parameter 1.

Finally, we note two cases of interest. First, we're sure about the value of θ for the second data set. To capture that, we let $n_2 \to \infty$, and the Kullback- Leibler divergence gets multiplied by n_1. Second, $n_1 = n_2$, as would happen when we're comparing US-US and CS-US intervals. In that case, the Kullback-Leibler divergence gets multiplied by $n_1/2$.