Published in final edited form as:

Behav Processes. 2013 May ; 95: 3–7. doi:10.1016/j.beproc.2013.01.005.

It's The Information!

Ryan D. Ward¹, C.R. Gallistel², and Peter D Balsam^{1,3}

¹Columbia University

²Rutgers University

³Barnard College

Abstract

Learning in conditioning protocols has long been thought to depend on temporal contiguity between the conditioned stimulus and the unconditioned stimulus. This conceptualization has led to a preponderance of associative models of conditioning. We suggest that trial-based associative models that posit contiguity as the primary principle underlying learning are flawed, and provide a brief review of an alternative, information theoretic approach to conditioning. The information that a CS conveys about the timing of the next US can be derived from the temporal parameters of a conditioning protocol. According to this view, a CS will support conditioned responding if, and only if, it reduces uncertainty about the timing of the next US.

Keywords

conditioning; timing; contiguity; associative models; information theory; informativeness; uncertainty

It is widely accepted that animals learn and encode the duration of events in conditioning protocols. Even though timing of events in conditioning protocols has long been demonstrated to have a profound impact on conditioned responding in a variety of paradigms (Blaisdell, Denniston, & Miller, 1998; Gallistel & Gibbon, 2000; Gibbon & Balsam, 1981; Miller & Barnet, 1993; Savastano & Miller, 1998; see Balsam, Drew, & Gallistel, 2010, for review), the importance of learning and encoding of temporal durations in theoretical treatments of conditioning has most often (with few exceptions) been relegated to the background in favor of conceptual and theoretical accounts which subscribe to the notion that contiguity between a CS and a US is the primary determinant of learning. According to this theoretical position, close temporal contiguity is the basis for the formation of associations between stimuli in a conditioning protocol.

Although the position that contiguity is the basic principle of learning has a long and storied history, empirical data have accumulated that are problematic for this view. A number of experiments demonstrated that repeated temporal contiguity between a candidate conditioned stimulus (CS) and a motivationally important event (US) was insufficient to establish conditioned responding. For example, Kamin (1967, 1969) showed that when rats were first conditioned with one CS, followed by conditioning to a second CS presented in

© 2013 Elsevier B.V. All rights reserved.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

compound with the first CS, conditioning to the new CS did not develop, notwithstanding its close temporal contiguity with the US. In one of the most striking of these experiments, Rescorla (1968) exposed different groups of rats to conditioning protocols which did not differ in the temporal pairing of CS and US, but differed in whether the US presentation was contingent on presentation of the CS. When the contingency between CS and US was made 0 by presenting US's during the intertrial interval (ITI) at the same rate they were presented during the CS, the rats did not develop a conditioned response to the CS. These experiments and the empirical demonstration of similar results (collectively called "cue competition" phenomena), such as overshadowing (Kamin, 1969; Reynolds, 1961) and relative validity (Wagner et al., 1968), indicated that the critical component of the relation between CS and US was not temporal contiguity, but rather the degree to which the US could be predicted given the occurrence of the CS. Put another way, a CS will support conditioned responding to the degree that it provides information about the occurrence of the next US. When Rescorla and Wagner (1972) introduced their canonical associative framework, however, they salvaged the contiguity-dependent view by parsing the continuous stream of stimuli and events experienced by organisms in a conditioning protocol into arbitrarily defined trials of an experimenter-specified length and assuming that the strength of conditioning depended on the prediction error of all cues present during a reinforced trial. The discrepancy between the summed "associative strengths" of all cues present and the asymptote determined the increment in associative value of individual cues. Thus, on a trial-by-trial basis the extent to which one cue has associative value limits the extent to which other cues can gain value. This model allows learning to be driven by contiguity but still permits cues to compete with each other for associative value. The general notion of contiguity-based increments and decrements in associative strength has been the conceptual foundation of the study of learning, including the search for the neurobiological basis of learning, for the last 50 years.

There are a growing number of empirical findings that pose problems for traditional associative models of conditioning (see Balsam, Drew, & Gallistel, 2010, for review), but for the purpose of brevity we focus here on one crucial stumbling block. The traditional results cited as evidence of the importance of contiguity in learning are data showing that as the interval between presentation of the CS and US increases, the strength of conditioned responding decreases. This effect has been widely demonstrated in a variety of preparations (e.g., Gibbon, Baldock, Locurto, Gold, & Terrace, 1977; Gormezano & Kehoe, 1981; Ost & Lauer, 1965; Reynolds, 1945; Smith, 1968; Stein, Sidman, & Brady, 1958; Wickens, Meyer, & Sullivan, 1961; Vandercar & Schneiderman, 1967) and cited as evidence for the critical role of contiguity in learning in conditioning protocols. However, it has been repeatedly demonstrated that the effect of increasing the CS-US interval on strength of conditioning depends crucially on the duration of the ITI. The decrement in conditioning that occurs with increasing CS-US interval is eliminated if the ITI is increased in proportion to the CS-US interval (see Gallistel and Gibbon, 2000, for review). Specifically, acquisition speed in conditioning protocols has been shown to be a function of the ratio of cycle time (C ; duration between successive US presentations) to trial time (T ; duration of conditioned stimulus presentation in delay conditioning; see Gibbon & Balsam, 1981; Gibbon et al., 1977). The number of trials to acquisition is generally similar with similar C/T ratios, regardless of the absolute values of C and T . Figure 1 shows the results from experiments (Gibbon et al., 1977) where pigeons were exposed to an autoshaping procedure and the effect of increasing CS-US interval on trials to acquisition was assessed. In groups for which the ITI was held constant, the trials to acquisition increased with increasing CS-US interval. However, in groups for which the ITI was increased proportionally to the CS-US interval, the number of trials to acquisition was constant, regardless of the CS-US interval. Thus, what matters in conditioning is not the absolute delay to reinforcement, as asserted by the traditional notion of contiguity, but rather the relative delay to reinforcement. This property

of conditioning, which has been demonstrated numerous times in a number of species and preparations, strongly suggests timescale invariance of the conditioning process.

The dire implications of timescale invariance for associative models of conditioning cannot be overstated (see Gallistel & Gibbon, 2001, for discussion). Simply put, current associative models have no way of dealing with timescale invariance because they are critically dependent on the notion of a trial of some specified length. Thus, intrinsic features of these models render them exquisitely sensitive to the absolute time scale of the conditioning protocol. As troublesome as timescale invariance is for formal models of associative learning given their trial-based structure, it is perhaps even more damning for the conceptual notion of contiguity as the fundamental principle of learning. If what matters in the learning and emergence of conditioned responding is the relative, not the absolute, delay to reinforcement, then what exactly constitutes a contiguous temporal pairing? What is the critical window of associability? There is no straightforward answer.

For this and other reasons we reject contiguity as the primary principle of learning, and offer an alternative conceptualization of the content and process of learning in conditioning protocols (Balsam & Gallistel, 2009; Balsam et al., 2010). In this paper, we review the primary conceptual and quantitative underpinnings of a model of conditioning based on an information theoretic analysis of the timing of events in conditioning protocols. A more complete treatment is given elsewhere (Balsam & Gallistel, 2009; Balsam et al., 2010).

The conceptual foundation of the information theoretic analysis is that in a conditioning protocol, events or cues (CS presentation) are informative to the extent that they tell the animal something it did not already know about the timing of the next US. The important information to be conveyed is how close (temporally speaking) the animal is to reinforcement. An informative CS is one that tells the animal that the US is relatively near. In other words, informative cues decrease the expected time to the next US. We assume, and extensive evidence has shown, that animals rapidly learn the duration of events in conditioning protocols (Balsam, Drew, & Yang, 2002; Drew, Zupan, Cooke, Couvillon, & Balsam, 2005; Kirkpatrick & Church, 2000a, 2000b; Oyama & Mauk, 2001). However, according to this view, the durations between events in a conditioning protocol do not determine the extent to which associations will form between the events. Rather, the durations between events *are* the content of learning in conditioning protocols. These learned durations form the basis for the computation of the expected time to reinforcement

Specifically, the underlying logic of the model is that a given cue (CS) will support conditioning to the extent that it reduces uncertainty about the timing of the next US. This is a conceptually intuitive idea, and allows a quantitative formalization made possible by Shannon's work in information theory (Shannon, 1948). According to Shannon's conceptualization, a signal is informative to the extent that it reduces the receiver's uncertainty about some stochastic aspect of the world. The information conveyed by a given signal can be quantified as the difference in the uncertainty in the presence of the signal and the uncertainty when the signal is ignored or never presented. Applied to conditioning protocols, this means that the information conveyed by a prospective CS is the difference in uncertainty about the timing of the next US in its presence and the uncertainty about the timing of the US in the context. Uncertainty is quantified as the entropies of probability distributions. The entropy measures the uncertainty associated with some random variable, in this case, the variables describe the distributions of CS-US and US-US intervals. Thus, by computing the difference between the US-US entropy and the CS-US entropy, we obtain a measure of the information (reduction in uncertainty) of the timing of the next US conveyed by presentation of the CS. In the basic conditioning protocol, USs are distributed according

to a random rate process, therefore the underlying distribution of inter US intervals is exponential. The entropy of an exponential distribution with an expectation μ is

$$\log_2 \mu + \log_2 \frac{e}{\Delta\tau} \quad \text{Equation 1}$$

where $\Delta\tau$ is the resolution with which time is measured by the animal. To compute the information that a given CS conveys about the timing of the next US, we subtract the entropy of the distribution after CS onset from the entropy of the distribution before CS onset. This gives the information conveyed by CS onset, H_{conv} , as follows

$$H = \left[\log_2 C + \log_2 \frac{e}{\Delta\tau} \right] - \left[\log_2 T + \log_2 \frac{e}{\Delta\tau} \right] = \log_2 C - \log_2 T = \log_2 \frac{C}{T} \quad \text{Equation 2}$$

Where C is the US-US interval and T is the CS-US interval. Notice that the term that depends on $\Delta\tau$, the precision with which the subject represents temporal intervals, disappears, leaving only the ratio between C and T as the determinant of the information conveyed by CS onset.

It should be clear from Equation 2 that the information conveyed by a CS depends on the relative temporal distance to reinforcement signaled by onset of the CS. As long as C is increased proportionally to T , the ratio of C to T does not change and neither does the amount of information conveyed by CS onset. Because the model predictions depend on the ratio of C and T , not their absolute durations, the model is timescale invariant. This is displayed graphically in Figure 2. The figure is drawn from the same data as Figure 1, but instead of plotting the trials to acquisition as a function of the duration of T , it plots the information conveyed by CS onset which can be derived from the temporal parameters of the protocol for each of the experimental groups. When the ITI is fixed, increasing the duration of T produces a decrease in the information that each CS onset provides about the timing of the next US. When the ITI is increased proportionally to T , thereby keeping the C/T ratio constant, the information conveyed by CS onset does not change when T is increased. These results mirror those from Figure 1, and provide a theoretical explanation for these data. Thus, trials to acquisition can be understood in terms of the information conveyed by the CS.

We call the component of information that varies with the C/T ratio informativeness, and it represents one component of the information potentially conveyed by CS presentation. If the distribution of both CS-US and US-US intervals is exponential, it is the only component. In the usual delay conditioning protocol, however, the duration of T is fixed. This introduces another component of the information about US timing conveyed by CS onset. This additional component does not depend on the temporal parameters of the protocol (i.e., on C and T), it depends only on how precisely a subject represents temporal intervals, as measured by the Weber fraction, that is by the coefficient of variation (σ/μ) of the distribution of some well-timed aspect of a conditioned response.

When the CS-US interval is fixed, the objective “distribution” of delays has 0 entropy, because there is no variation in the delay, hence no uncertainty about when the US will occur once the CS has come on. The objective distribution has 0 entropy regardless of the duration of the fixed delay. In that case, the uncertainty about when to expect the US after the onset of the CS arises from the scalar imprecision of the brain’s representation of magnitudes of all kinds, including temporal magnitudes (interval durations). Many industrially important physical quantities (e.g., electrical resistance) are specified by the manufacturer only to $\pm$ some percentage of their nominal value, and so it is with the brain’s specification of experienced magnitudes. This long established fact is called Weber’s

law. Thus, the uncertainty about when to expect the US once the CS has come on in a delay-conditioning paradigm is a subjective uncertainty rather than an objective uncertainty. That is why the measure of this uncertainty depends on a parameter of the subject (the temporal Weber fraction) rather than on one or more parameter(s) of the protocol.

We assume that the uncertainty about when to expect the US following the onset of the CS in a delay conditioning protocol is the entropy of a normal distribution with a standard deviation proportional to the mean, that is, a normal distribution with scalar variability. The entropy of a normal distribution is:

$$H = \frac{1}{2} \log_2 \left[2\pi e \frac{\sigma^2}{(\Delta\tau)^2} \right].$$

The difference, H_{conv} , between this entropy and the basal entropy (the measure of the uncertainty before the onset of the CS) is the information conveyed by CS onset:

$$\begin{aligned} H_{\text{conv}} &= (\log_2 e + \log_2 C - \log_2 \Delta\tau) \\ &\quad - \left(\frac{1}{2} \log_2 2\pi e + \log_2 T + \log_2 w - \log_2 \Delta\tau \right) \\ &= \log_2 e + \log_2 C - \log_2 \Delta\tau \\ &\quad - \frac{1}{2} \log_2 2\pi \\ &\quad - \frac{1}{2} \log_2 e \\ &\quad - \log_2 T - \log_2 w + \log_2 \Delta\tau \\ &= \frac{1}{2} (\log_2 e \\ &\quad - \log_2 2\pi) \\ &\quad + \log_2 C - \log_2 T - \log_2 w = \log_2 \frac{C}{T} + \frac{1}{2} \log_2 \frac{e}{2\pi} - \log w \end{aligned} \quad \text{Equation 3}$$

Notice that the term that depends on the precision no longer disappears, which is why the Weber fraction, w , is present. It might be thought that this is because we take account of the “smudging” effect of the subjective imprecision in representing temporal intervals when we write the expression for the subject’s uncertainty after CS onset (2nd term in first line of Equation 3) but we fail to do so when we write the expression for the subject’s basal uncertainty (1st term in first line of Equation 3). Put another way, does not the fuzziness with which a subject represents temporal intervals increase the basal uncertainty as well as the uncertainty that prevails after CS onset? The answer is, no, it does not. Mathematically, this is because convolving the exponential distribution with a Gaussian yields the exponential. Intuitively, this must be the case, because, for a given expectation, the exponential distribution is the maximum entropy distribution. It is impossible to make it more random than it already is; it is unsmudgeable.

The Weber fraction has been shown to be around 0.16 in pigeons, rats, and mice (Gallistel, King, & McDonald, 2004). Evaluating Equation 3 with this value for w and comparing the result to that obtained from evaluating Equation 2, we find that fixing the CS-US interval adds about 2 bits of information (see Balsam et al., 2010). The amount of this addition does not depend on value of T , the expected delay of reinforcement; it depends only on the value of w , the Weber fraction, which measures the precision with which subjects represent the durations of intervals. Surprisingly, we have recently shown that this additional information,

which is equivalent to a 4-fold change in C/T , does not affect acquisition speed (Ward et al., 2012). Thus, informativeness—the factor by which the onset of the CS shortens the expected delay of reinforcement—is the critical quantity modulating acquisition speed, not the total amount of information conveyed by the onset of the CS. When conditioning protocols are analyzed according to this framework, and predictions made in terms of the information conveyed by cue presentation about the timing of the US, the information theoretic framework can account for a broad array of Pavlovian conditioning results, including those that are troublesome for contiguity-based approaches (see Balsam & Gallistel, 2009; Balsam et al., 2010, for review and discussion).

The differing effects from the two different sources of temporal uncertainty—the objective uncertainty inherent in a randomly varying delay and the subjective uncertainty arising from limited precision in the measurement of those delays—raises the question whether subjects can distinguish these uncertainties. It has generally been assumed that they cannot. It has been assumed the purely subjective uncertainty arises from random variation in the neural mechanism that measures intervals, that is, from noisy interval timing (Gibbon 1977; Killeen and Fetterman 1988; Gallistel and Gibbon 2000). If that were so, then it is hard to see how the brain could distinguish between variability in the intervals timed and variability in the timer; the variance in the recorded interval durations would have to be the sum of the external variance and the variance coming from the measuring mechanism.

There are, however, other possible explanations for the subjective imprecision. Some level of imprecision is inherent in any physically realized symbolic representation of continuous magnitude. For example, in choosing the number of digits used in specifying a magnitude, one makes an implicit claim about the precision with which the magnitude is known or need be known for present purpose. Any physically realized mechanism for representing quantities that vary over many orders of magnitude, as do temporal intervals, must employ some form of autoscaling (Gallistel 2011), and this must scale the precision of the representation to the magnitude (Weber's law). In other words, Weber's law may reflect a well-engineered feature of the brain's mechanism for representing magnitude rather than a bug arising from an inability to suppress internal noise. In that case, it may be possible for the brain to distinguish the objective uncertainty that arises from random variability in experienced intervals from the uncertainty inherent in the limited precision with which it chooses to represent those intervals.

Results obtained by Aaron Kheifets in the Gallistel lab, which are now being prepared for publication, imply that the mouse does make this distinction. Kheifets added objective noise to the temporal intervals used in the “switch” paradigm first developed by Fetterman and Killeen (1995) and further developed by Balci, Freestone and Gallistel (2009) and Kheifets and Gallistel (2012) for use in the analysis of interval timing and risk estimation in the mouse. Instead of increasing the variability in the timing of the switches, as would be expected if objective variability were confounded with subjective variability, making the switch task harder by adding this objective variability reduced quite substantially the variability in the timing of the switches. That is, the mice rose to the challenge.

The now abundant evidence that the brains of non-human animals encode temporal intervals ranging over many orders of magnitude and rely on the information thus preserved to inform subsequent behavior raises the obvious question of how it does so. The question has two parts: How does the brain measure temporal intervals? And, how does it preserve a record of its measurements? A variety of models have been suggested in answer to the first question, some motivated by neurobiological considerations and some not (for reviews, see Meck 2003; Buhusi and Meck 2005, Chap 15; Gallistel and King 2009). Whether motivated by neurobiological plausibility considerations or not, almost all models have assumed the use of

interval timing mechanisms to measure the experienced intervals. There is a conceptual problem with this assumption in that it is unclear how the process gets off the ground. The first time an event occurs that marks the onset of a behaviorally important interval, there is no knowledge of whether to start an interval timer nor how many to start, because there is no knowledge of what may or may not follow that event. The possibilities are infinite; therefore the system should start an infinite number of timers every time a novel event occurs. But that is clearly impossible. However if the system does not time intervals on first encounter, then there is no record of them, in which case every subsequent encounter is effectively a first encounter. No one has suggested a solution to this problem.

An alternative method of measuring interval durations, which does not raise this “getting-off-the-ground” problem, is to record times of event occurrence and obtain the intervals retrospectively by subtracting the time of occurrence of the earlier event from the time of occurrence of the later event. Times of occurrence may be obtained from the phases of the many different neural oscillators known to be present not only in neural tissue but also in non-neural tissues. Circadian and infradian cycles are present even in bacteria and plants, so plausible signal sources are at hand. What is hard to envisage in our current state of ignorance is the mechanism(s) by which the phases of these oscillations may be read off and recorded in memory.

This difficulty reflects the more general problem that current neurobiological conceptions of memory say nothing very specific about how the posited memory mechanisms store the information (the facts) gleaned from experience in computationally accessible form (Gallistel & King 2009; Gallistel & Matzel 2013). Our current inability to specify a neural mechanism by which facts gleaned from experience may be stored is evident in the way in which the neurobiologically motivated theories of timing address the second question: how does the brain preserve the records of previous measurements (in some useable form)? The neurobiologically motivated theories discussed above attempt only to explain the fact that conditioned behavior is timed. They do this by assuming that individual neurons or constellations of neurons that become active at the appropriate time become associated with the anticipatory response. In other words, they are S-R theories. Like all such theories, they are a-representational or anti-representational: the brain responds as if it knew the interval, but it does not have that information stored in symbolic form, which is to say, in a form that makes it useable in computational operations. Insofar as information about the durations of the experienced intervals can be said to reside somewhere in these conditioned reflex arcs, it resides in the intrinsic dynamics of the neurons on the afferent side of the arc (Karmarkar & Buonomano, 2007). There is no explanation of how any other part of the brain could have access to those dynamics. Thus, these models do not attempt to explain the fact that acquired temporal information informs behavior in many other ways. They do not, for example, attempt to explain how it is possible for the speed of acquisition to depend not on any fixed interval but rather on the ratio of the expectations of two randomly varying intervals (the US-US interval and the CS-US interval). To explain that, one must specify the mechanism by which the measurements of the durations of individual intervals are stored. The specification of the neurobiological mechanism by which these durations are encoded must make it clear how the brain could then access these stored measurements in order to compute the mean of a population of measurements and how it could take the ratio of two such means.

The information theoretic account summarized here is an attempt to specify the content of learning that governs speed of acquisition. In this regard, it differs from traditional associative theories, which typically make predictions about post-acquisition differences in learning, rather than speed of acquisition, or as we have termed it here, associability. Thus, comparison of the models is difficult given the differences in dependent measures.

In conclusion, it has been repeatedly demonstrated that contiguity is neither necessary nor sufficient to support conditioned responding. Given all these considerations, we think it difficult to sustain temporal contiguity as the fundamental principle of learning. We suggest as an alternative an analysis of how organisms extract meaningful regularities from the complex temporal stream of experience, and how these regularities, once extracted, guide responding. We further suggest that a focus on defining the informational content of conditioning protocols provides a start to such an analysis by specifying the content of learning that guides the emergence of conditioned responding.

References

- Balci F, Freestone D, et al. Risk assessment in man and mouse. *Proceedings of the National Academy of Science U S A*. 2009; 106:2459–2463.
- Balsam PD, Gallistel CR. Temporal maps and informativeness in associative learning. *Trends in Neurosciences*. 2009; 32:73–78. [PubMed: 19136158]
- Balsam PD, Drew MR, Gallistel CR. Time and associative learning. *Comparative Cognition & Behavior Reviews*. 2010; 5:1–22. Retrieved from <http://psyc.queensu.ca/ccbr/index.html>. [PubMed: 21359131]
- Balsam PD, Drew MR, Yang C. Timing at the start of associative learning. *Learning and Motivation*. 2002; 33:141–155.
- Balsam PD, Fairhurst S, Gallistel CR. Pavlovian contingencies and temporal information. *Journal of Experimental Psychology: Animal Behavior Processes*. 2006; 32:284–294. [PubMed: 16834495]
- Blaisdell AP, Denniston JC, Miller RR. Temporal encoding as a determinant of overshadowing. *Journal of Experimental Psychology: Animal Behavior Processes*. 1998; 24:72–83. [PubMed: 9438967]
- Buhusi CV, Meck WH. What makes us tick? functional and neural mechanisms of interval timing. *Nature Reviews Neuroscience*. 2005; 6:755–765.
- Drew MR, Zupan B, Cooke A, Couvillon PA, Balsam PD. Temporal control of conditioned responding in goldfish. *Journal of Experimental Psychology: Animal Behavior Processes*. 2005; 31:31–39. [PubMed: 15656725]
- Fetterman JG, Killeen PR. Categorical scaling of time: Implications for clock-counter models. *Journal of Experimental Psychology: Animal Behavior Processes*. 1995; 21:43–63. [PubMed: 7844506]
- Gallistel CR. Mental magnitudes. In: Dehaene, S.; Brannon, L., editors. *Space, Time and Number in the Brain: Searching for the foundations of mathematical thought*. New York: Elsevier; 2011. p. 3–12.
- Gallistel CR, Gibbon JG. Time, rate, and conditioning. *Psychological Review*. 2000; 107:289–344. [PubMed: 10789198]
- Gallistel CR, Gibbon JG. Computational versus associative models of simple conditioning. *Current Directions in Psychological Science*. 2001; 10:146–150.
- Gallistel CR, King A, McDonald R. Sources of variability and systematic error in mouse timing behavior. *Journal of Experimental Psychology: Animal Behavior Processes*. 2004; 30:3–16. [PubMed: 14709111]
- Gallistel CR.; King, AP. *Memory and the computational brain: Why cognitive science will transform neuroscience*. New York: Wiley/Blackwell; 2009.
- Gallistel CR, Matzel LD. The neuroscience of learning: Beyond the Hebbian Synapse. *Annual Review of Psychology*. 2013; 64:169–200.
- Gibbon J. Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*. 1977; 84:279–325.
- Gibbon J, Baldock MD, Locurto CM, Gold L, Terrace HS. Trial and intertrial durations in autoshaping. *Journal of Experimental Psychology: Animal Behavior Processes*. 1977; 3:264–284.
- Gibbon, J.; Balsam, PD. The spread of association in time. In: Locurto, C.; Terrace, HS.; Gibbon, J., editors. *Autoshaping and conditioning theory*. New York: Academic Press; 1981. p. 219–253.

- Gormezano, I.; Kehoe, E.J. Classical conditioning and the law of contiguity. In: Hazern, P.; Zeiler, MD., editors. Predictability, correlation and contiguity. New York: Wiley; 1981. p. 1-45.
- Kamin, L.J. Attention-like processes in classical conditioning. In: Jones, MR., editor. Miami symposium on the prediction of behavior: Aversive stimulation. Coral Gables, FL: University of Miami Press; 1967. p. 9-31.
- Kamin, L.J. Predictability, surprise, attention, and conditioning. In: Campbell, BA.; Church, RM., editors. Punishment and aversive behavior. New York: Appleton-Century-Crofts; 1969a. p. 276-296.
- Karmarkar UR, Buonomano DV. Timing in the absence of clocks: encoding time in neural network states. *Neuron*. 2007; 53:427–438. [PubMed: 17270738]
- Kheifets A, Gallistel CR. Mice take calculated risks. *Proceedings of the National Academy of Sciences USA*. 2012; 109:8776–8779.
- Killeen PR, Fetterman JG. A behavioral theory of timing. *Psychological Review*. 1988; 94:455–468. [PubMed: 3317471]
- Kirkpatrick K, Church RM. Stimulus and temporal cues in classical conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*. 2000a; 26:206–219. [PubMed: 10782435]
- Kirkpatrick K, Church RM. Independent effects of stimulus and cycle duration in conditioning: The role of timing processes. *Animal Learning and Behavior*. 2000b; 28:373–388.
- Meck, WH., editor. Functional and neural mechanisms of interval timing. New York: CRC Press LLC; 2003.
- Miller RR, Barnet RC. The role of time in elementary associations. *Current Directions in Psychological Science*. 1993; 2:106–111.
- Ost, JWP.; Lauer, DW. Some investigations of salivary conditioning in the dog. In: Prokasy, WF., editor. Classical conditioning. New York: Appleton-Century-Crofts; 1965.
- Ohyama T, Mauk MD. Latent acquisition of timed responses in cerebellar cortex. *Journal of Neuroscience*. 2001; 21:682–690. [PubMed: 11160447]
- Rescorla RA. Probability of shock in the presence and absence of CS in fear conditioning. *Journal of Comparative and Physiological Psychology*. 1968; 66:1–5. [PubMed: 5672628]
- Rescorla, RA.; Wagner, AR. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In: Black, AH.; Prokasy, WF., editors. Classical Conditioning II: Current Research and Theory. New York: Appleton-Century-Crofts; 1972. p. 64-99.
- Reynolds B. The acquisition of a trace conditioned response as a function of the magnitude of the stimulus trace. *Journal of Experimental Psychology: Animal Behavior Processes*. 1945; 35:15–30.
- Reynolds GS. Attention in the pigeon. *Journal of the Experimental Analysis of Behavior*. 1961; 4:203–208. [PubMed: 13741095]
- Savastano HI, Miller RR. Time as content in Pavlovian conditioning. *Behavioural Processes*. 1998; 44:147–162.
- Shannon CE. A mathematical theory of communication. *Bell Systems Technical Journal*. 1948; 27:379–423. 623–656.
- Smith MC. CS-US interval and US intensity in classical conditioning of the rabbit's nictitating membrane response. *Journal of Comparative and Physiological Psychology*. 1968; 66:679–687. [PubMed: 5721496]
- Stein L, Sidman M, Brady JV. Some effects of two temporal variables on conditioned suppression. *Journal of the Experimental Analysis of Behavior*. 1958; 1:153–162. [PubMed: 16811211]
- Vandercar DH, Schneiderman N. Interstimulus interval functions in different response systems during classical discrimination conditioning of rabbits. *Psychonomic Science*. 1967; 9:9–10.
- Wagner AR, Logan FA, Haberlandt K, Price T. Stimulus selection in animal discrimination learning. *Journal of Experimental Psychology: Learning, Memory & Cognition*. 1968; 76:171–180.
- Ward RD, Gallistel CR, Jensen G, Richards VL, Fairhurst S, Balsam PD. Conditioned stimulus informativeness governs conditioned stimulus-unconditioned stimulus associability. *Journal of Experimental Psychology: Animal Behavior Processes*. 2012; 38:217–232. [PubMed: 22468633]

Wickens DD, Meyer PM, Sullivan SN. Classical GSR conditioning, conditioned discrimination, and interstimulus intervals in cats. *Journal of Comparative and Physiological Psychology*. 1961; 54:572–576. [PubMed: 14006713]

Highlights

- Temporal contiguity between a conditioned stimulus and an unconditioned stimulus is insufficient as an explanation for learning in conditioning protocols
- An analysis of conditioning protocols based on information theory provides an alternative to contiguity-based accounts
- An informational analysis states that a CS will support conditioned responding if it reduces uncertainty about the timing of the next US

Figure 1.

Acquisition speed as a function of trial CS duration. Different groups of pigeons were exposed to an autoshaping protocol with fixed delays that ranged from 4 s to 32 s. For some groups, the duration of the intertrial interval was kept constant (filled circles). For these groups, the number of trials to acquisition increased with increased CS duration. In other groups (filled squares) the ratio of the intertrial to trial CS duration was kept constant. In these groups, speed of acquisition remained constant regardless of the duration of the trial CS. After Balsam et al. (2010). Original data from Gibbon et al., (1977).

Figure 2.

CS informativeness as a function of trial CS duration. Bits of information per CS presentation across the different groups displayed in Figure 1. Bits of information was calculated according to Equation 2. For groups in which the duration of the intertrial interval was constant (filled circles), bits of information conveyed by each CS presentation decrease as the duration of the trial CS increases. For groups in which the ratio of the intertrial interval to trial CS duration was kept constant (filled squares), bits of information conveyed by each CS presentation remains constant across increasing CS duration.