

Minimum description length model selection in associative learning

C Randy Gallistel¹ and Jason T Wilkes²



Two principles of information theory — maximum entropy and minimum description length — motivate a computational model of associative learning that explains assignment of credit, response timing, and the parametric invariances. The maximum entropy principle gives rise to two distributions — the exponential and the evitable Gaussian — which naturally lend themselves to inference involving each of the two fundamental classes of predictors — enduring states and point events. These distributions are the ‘atoms’ from which more complex representations are built. The representation that is ‘learned’ is determined by the principle of minimum-description-length. In this theory, learning is a synonym for data compression: The representation of its experience that the animal learns is the representation that best allows the data of experience to be compressed.

Addresses

¹Rutgers Center for Cognitive Science, 152 Frelinghuysen Rd, Piscataway, NJ 08854-80210, United States

²Department of Psychology, UCSB, Santa Barbara, CA, United States

Corresponding author: Gallistel, C Randy (galliste@rucss.rutgers.edu)

Current Opinion in Behavioral Sciences 2016, 11:8–13

This review comes from a themed issue on **Computational modeling**

Edited by **Peter Dayan** and **Daniel Durstewitz**

<http://dx.doi.org/10.1016/j.cobeha.2016.02.025>

2352-1546/Published by Elsevier Ltd.

Introduction

Associative learning, as studied in Pavlovian and operant conditioning paradigms, is an extensively studied and frequently modeled cognitive process. Most computational models of it are associative models; they assume that the learning consists of the alteration of connection strengths between elements of the mind (nodes) or brain (neurons). The assumption that changes in associative strength are realized in the brain by changes in synaptic conductances is explicit in most research aimed at establishing the neurobiological basis of learning and memory. Thus, the question whether the associative theories can be made to yield an account of well established experimental results — and if not, what alternative theories

can — is important to progress in determining the neurobiological bases of learning and memory.

Though conditioning phenomena are commonly referred to as ‘associative’ learning, associative theories have failed for decades to capture many well-established experimental results. No associative theory known to us addresses all three of the following well-established bodies of experimental results: (1) The assignment of credit (aka cue competition). (2) The timing of the conditioned response. (3) The three parametric invariances: (i) Time-scale-invariance: trials to acquisition is proportionate to the ratio of total CS time to total training time [1,2*]. This means, among things, that for fixed total training time, the number of CS-US pairings is irrelevant [3*]. (ii) Reinforced trials to acquisition is invariant under partial reinforcement [4]. (iii) Omitted reinforcements to extinction is invariant under partial reinforcement during training [4].

For example, the models of Rescorla and Wagner [6], Mackintosh [7], and Pearce and Hall [8] address cue competition but not response timing nor the parametric invariances, while associative models that address response timing [9,10] do not address credit assignment nor the parametric invariances. Some temporal difference learning models have addressed — to some extent — both response timing and credit assignment [11,12], but they do not address the parametric invariances.

A step in the right direction has been provided by non-associative models of ‘associative’ learning [13,14,15*]. These have the dual virtues of (i) framing behavior in conditioning paradigms as the product of normative statistical inference, and (ii) allowing the complexity of the statistical model formed by the brain to depend on the data. However, the support for the distributions in these normative-statistical-inference theories is trial types. To the animal, there is no such thing as a trial.

We have recently elaborated a non-associative, causal, algorithmic, real-time theory of associative learning [16*] that addresses all three bodies of evidence. It is founded on two principles of information theory: maximum entropy and minimum description length (MDL).

Core assumptions

We assume that the brain represents candidate predictors — possible causes of some event of interest — in one of two basic forms: all predictors are either *states* (such as a light *being* on or off) or *point events* (such as a light *turning*

on or off). By definition, states are perceived to have duration — a temporal interior — while point events are perceived to be located at a point in time, with no measurable duration. States are simply continuous intervals, whose onsets and offsets are themselves point events. They have, therefore, the potential to convey *rate* information (e.g., ‘the average rate of a shock is 3 times per minute *during* state S’). Point events have the potential to convey *timing* information (e.g., ‘a shock tends to occur approximately 10 seconds after the *offset* of state S’).

The principle of maximum entropy [41,42] is a powerful method for translating *qualitative* prior information into *quantitative* form in a way that embodies the information fed into it, but no more. Given minimal assumptions, maximum entropy effectively builds a model for us out of a data set. For example, suppose we have a set of data points generated by a distribution of unknown form. We can then make different (qualitative) assumptions about the distribution that generated that data, and examine the resulting (quantitative) models. If we choose, as our set of background assumptions, a statement such as, ‘The mean is roughly m . That’s all I know,’ then the maximum entropy formalism tells us that our state of knowledge is most honestly represented by the exponential distribution. If we choose as our set of background assumptions, the statement, ‘The mean is roughly m . The variance is roughly v . That’s all I know,’ then the maximum entropy formalism tells us that our state of knowledge is most honestly represented by the Gaussian distribution.

The seemingly independent narratives of the two paragraphs above turn out to be related in a surprisingly elegant manner. As it happens, the ‘first’ maximum entropy distribution — the exponential — is naturally suited to inference involving *states*, while the ‘second’ maximum entropy distribution — the Gaussian — is naturally suited to inference involving *point events*. This elegant relationship forms the foundation of our theory of associative learning.

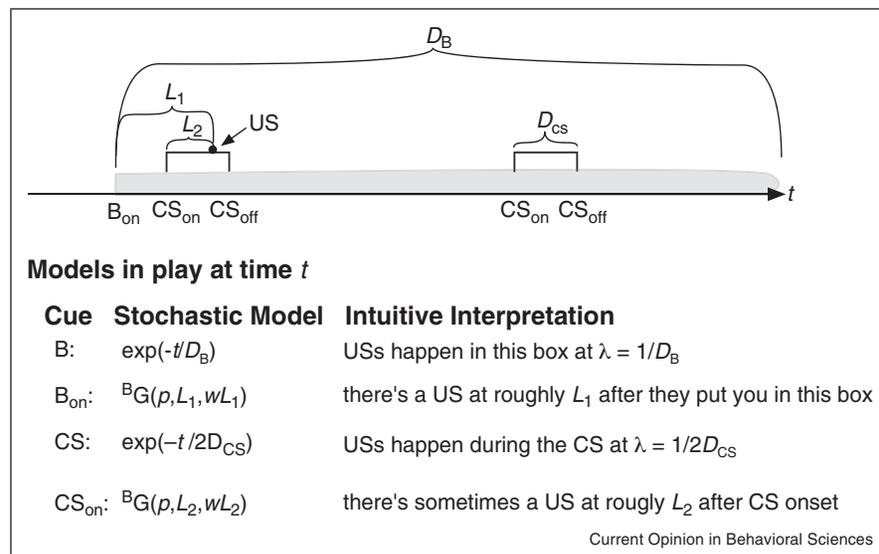
One moment known (m) → exponential → states → rates.

Two moments known (m and v) → Gaussian → point events → time.

These two cognitive building blocks — states and point events; one dimensional intervals and zero dimensional points — are the atoms from which candidate representations of temporal experience are built. Given a conditioning protocol with a single CS, we have one state cue (the state ‘CS = on’) and two point cues (the events ‘CS onset’ and ‘CS offset’). The experimental chamber itself constitutes an additional state cue, while the moments at which the animal enters and leaves it constitute additional point cues. Thus, even in a simple experiment involving a single CS, we already find ourselves with a nontrivial number of candidate predictors, and a correspondingly nontrivial number of candidate *hypotheses* as to what might be causing the USs the animal experiences in the chamber.

An animal is removed from its cage and placed in an experimental chamber. One minute later, a light turns on,

Figure 1



The time line at top shows the first two trials (CS presentations, white rectangles) of one of Rescorla's [17] protocols. The subject is placed in the box at B_{on} (gray background). The US occurs at latency L_1 if its latency is measured from B_{on} or L_2 if measured from CS_{on}. The state cues are the Background (gray) and the CS (white). The appropriate stochastic model is ambiguous at time t ; any of the 4 listed are viable.

and stays on for one minute (see [Figure 1](#)). Forty seconds after the onset of the light, the animal experiences a single shock. Five minutes later, the light appears again for one minute, but with no shock. Given this limited information, how does it represent its situation in order to effectively predict the future? Was the shock caused by the onset of the light, with a 40 s delay? Or was it caused by the state ‘Light = on,’ with a random rate of roughly one shock per 2 min? Or was it caused by the state of being in the experimental chamber, with a rate of roughly one shock every ten minutes? Or perhaps the point event of being dropped into the experimental chamber itself is to blame. Perhaps every such placement will be followed by a single shock after a span of roughly 100 s. Our own intuition is not silent here. Faced with the above thought experiment, we readily find ourselves entertaining a set of *candidate hypotheses*, and we cannot help but find some of them more appealing than others.

Our theory is that the brain parses temporal experience into states and point events, and uses these primitives to arrive at a set of candidate hypotheses not unlike our intuitions ([Figure 1](#)). The manner in which a particular hypothesis is selected — that is, how a particular representation is learned in our theory — is determined by an algorithmic implementation of the *principle of minimum description length* [[43,44](#)].

MDL is a principle for stochastic model selection based on a simple premise: any statistical regularity in a set of data can be used to compress those data. This follows from Shannon’s coding theorem, which establishes a one-to-one mapping between the probability of a datum (its relative frequency in the data set) and the relative length of the code for that datum in a maximally efficient encoding of the data. Better models assign higher probabilities, so they enable more efficient encodings (shorter codes). However, to make the compressed data decodable, the stochastic model that dictated the relative code lengths must itself be encoded. Thus, the total memory cost (total bits required) is the sum of the bits required for the compressed encoding of the data plus the bits required to encode the stochastic model that mediated the compression. Taking into account the cost of storing the stochastic model along with the data forestalls the use of excessively complex models. MDL provides a mathematical codification of the folk concept of parsimony, in a manner that allows a simultaneous solution of two fundamental computational problems: selection of a predictive model and efficiency of information storage.

The evitable Gaussian

The ordinary Gaussian distribution does not allow for the possibility that the predicted event will fail to happen at anywhere near the predicted time, but predicted events do sometimes fail to happen, both in life and in conditioning experiments. To allow for this, we introduce the evitable

Gaussian form, which we call the Bernoulli Gauss, and which we denote by ${}^B\text{G}(t|t_0, p, \mu, \sigma)$. The cumulative ${}^B\text{G}$ distribution function specifies the cumulative probability of an anticipated event as a function of, t , for an event that occurs with probability, p , following a point cue at t_0 , with an expected latency, μ , and an expected temporal dispersion measured by σ (see [Figure 2](#)).

The vector equation is:

$${}^B\text{G}(t|t_0, p, \mu, \sigma) = \left\{ p\Phi\left(\frac{t-t_0-\mu}{\sigma}\right) \quad (1-p)\left(1-\Phi\left(\frac{t-t_0-\mu}{\sigma}\right)\right) \right\}$$

for $t \geq t_0$; 0 otherwise; where t_0 is the time of occurrence of the point cue from which the latency is measured, $0 \leq p \leq 1$, $w\mu \leq \sigma \leq .5\mu$, where w is the Weber Fraction for duration, whose measured values generally fall in the range 0.08–0.35, and Φ is the cumulative Gaussian distribution function.

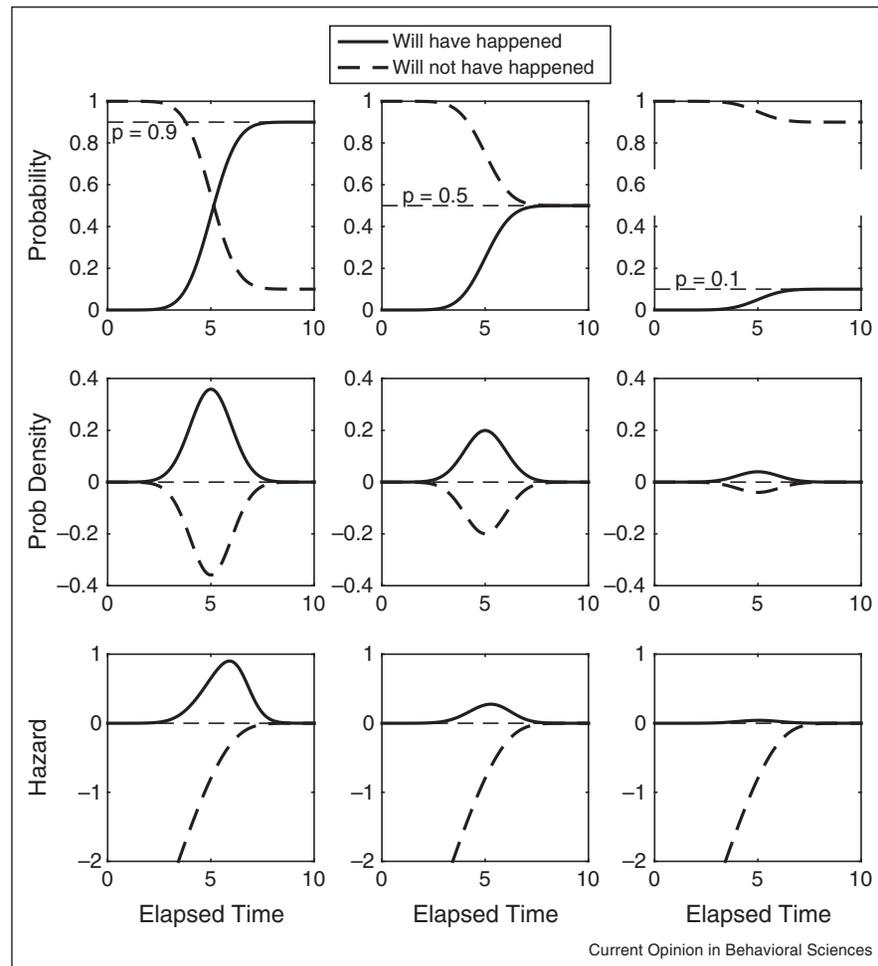
Cue competition and response timing

The issues of cue competition and response timing are together evident when considering [Figure 1](#), which shows the first two CS presentations in an experimental protocol like Rescorla’s [[17](#)] truly-random-control experiment. Beneath the time-line for the protocol are 4 plausible stochastic models for encoding the US (indicated by the dot) and predicting the wait times for future USs. In this review, we use US (unconditioned stimulus aka reinforcement) as a generic term for the event on the right-hand side of an associative bond in associative theories of associative learning. We use, CS, short for conditioned stimulus, for the predictive cue, whether state or point, the element on the left of an associative bond.

In our computational model, the 4 simplest predictive models in play at time, t , in [Figure 1](#) are:

- (1) USs occur at random intervals in the test box (the context, indicated by the gray in [Figure 1](#)). This is the ‘truly random control’ protocol [[17](#)]. Under these conditions, the Background:E model (the model in which the background predicts the parameter of an exponential distribution) best compresses the data (the experienced inter-event intervals). Its flat hazard function predicts steady responding, as is in fact observed in random time and in variable-reinforcement-interval protocols [[18](#),[19](#)].
- (2) USs occur at random but only during the CSs. This is the ‘contingent’ protocol [[17](#)]. Under these conditions, the CS:E model best compresses the data. An appropriately timed conditioned response is freezing during the CSs, which is what is observed in subjects run in the CS-contingent protocol [[17](#)].
- (3) Placement in the box (B_{on}) predicts a single US at latency L_1 , with probability p_1 ; during the rest of the session, there are no further USs. This is the

Figure 2



(Top) The cumulative ${}^B G$ distribution function maps an elapsing latency, $L = t - t_0$, into a 2D vector, the first element of which (solid curve) specifies the probability that the event will have happened, the second element of which specifies the complementary probability that it will not yet have happened (dashed curve). **(Middle)** The 2 elements of the probability density vector, which is the derivative of the cumulative probability vector. **(Bottom)** The two elements of the hazard vector. The hazard at any time t is the probability density divided by the survival probability. Intuitively, the positive element of the hazard vector is the probability that the event will happen in the next moment given that it has not yet happened. The columns are for values of p ranging from .9 (leftmost column) to .5 (middle) to .1 (rightmost).

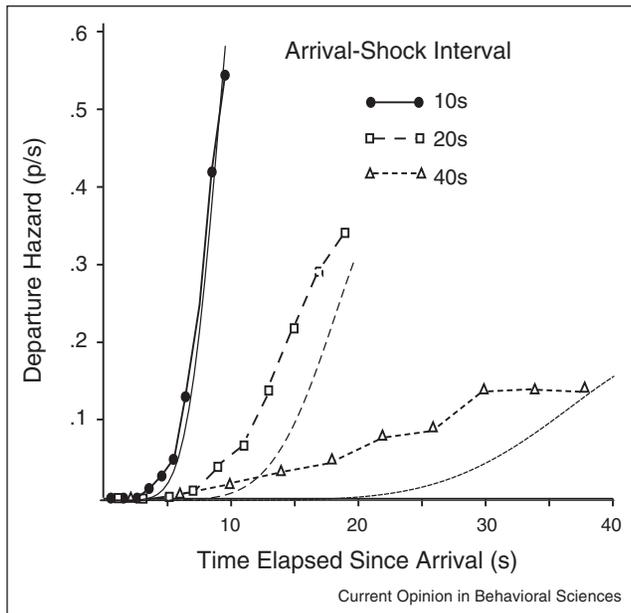
immediate-shock deficit protocol [20,21]. Under these conditions, the $B_{on} : {}^B G(p, \mu, \sigma)$ model (the model in which the onset of the background state predicts the time and probability of a US) best compresses the shock-latency data. Its rising and falling hazard function (Figure 2, bottom, black curves) predicts freezing localized in an interval surrounding the latency of the short latency of the single shock [1,22]. In shuttle-box shock-avoidance protocols, wherein a shock occurs at a fixed latency following the subject's arrival in either of the two boxes between which it shuttles to avoid those shocks [23], this hazard function predicts a departure when some proportion of the shock latency has elapsed (Figure 3).

- (4) CS onsets predict a single US at latency L_2 after CS_{on} , with probability p_2 . This is a 'delay conditioning'

protocol, the most common protocol in Pavlovian conditioning experiments. In such a protocol, the $CS_{on} : {}^B G(p, \mu, \sigma)$ model best compresses the latency data. Its hazard function predicts an increased probability of anticipatory responding as the expected time of the US approaches ([24], see their Figure 3). Our theory further predicts the cessation of anticipatory responding observed in the so-called peak procedure, when the CS extends well beyond the expected moment or reinforcement without a reinforcement [25,26] — see Figure 4.

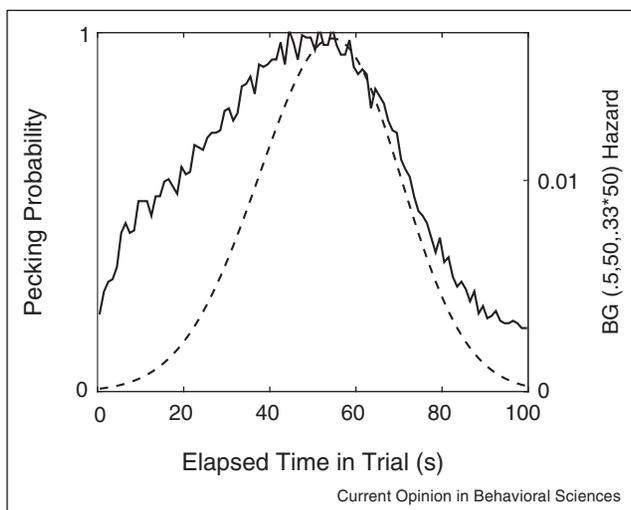
In our computational theory of the learning process closed-form formulae give the net cost of encoding the inter-reinforcement intervals so far experienced at each moment in time in a given protocol. Computational

Figure 3



The shuttle-box departure hazard (probability of leaving a box in the next second given that the subject has not yet left it) as a function of the time elapsed since the subject's arrival in the box, for varying fixed arrival-shock latencies (10, 20 and 40 s) (curves with data points retraced from [23] Figure 1). The curves without data points are the computational model's hazard functions for this protocol; they roughly predict the observed response timing. Better prediction is achieved by introducing trial-to-trial noise in the threshold for a departure.

Figure 4



Jagged curve is probability of pecking as function of time elapsed in a peak trial (data from one pigeon in laboratory of the late John Gibbon). Smooth dashed curve is the $BG(p = .5, L = 50s, \sigma = .33L)$ hazard function. Again, better prediction may be achieved by introducing noisy decision thresholds for the start and stop of responding.

implementation (in Python and in Matlab) confirms that the real-time application of the minimum-description length principle to these protocols yields the classic asymptotic cue competition results [17,27,28], including the retroactive blocking and unblocking results [29–32]. For reviews of these results, see [1,33,34] or a textbook on animal learning written after 1980. For a fuller explanation of why the minimum description length principle asymptotically picks out the solution that rats, pigeons, mice and humans asymptotically pick out, see [16*]. The basic reason is simple: The cost of encoding the data and the model-complexity cost both increase with the size of the data set. However, the first cost grows linearly, while the second grows only logarithmically. If, in the long run, a more complex model, with its higher complexity surcharge, captures structure that is really there in the sequence of intervals, then it reduces the linear rate at which the cost of encoding those intervals grows, relative to the rate from a model that fails to capture this structure. A more complex model prevails over a simpler model in the long run only if superior data compression compensates for increased complexity cost.

Neurobiological implications

The broad success of this non-associative model relative to associative models of associative learning suggests that the primitives in associative learning are stochastic models and data encoded in a symbolic read-write memory. By preserving accumulated information in retrievable form, read-write memories liberate computation from the tyranny of the current moment [35], enabling the brain to recode past information in the light of subsequent information [36]. They play no role in associative theories of associative learning, because plastic synapses (aka associative bonds or connection weights), which are the primitives in associative theories, do not serve as symbols; they do not refer and their values are not accessible to computation in the way in which bit patterns in RAM or codon sequences in a gene are accessible to computation [35,37*]. The recent finding that the memory for the CS-US interval in eyeblink conditioning resides at the molecular level inside the cerebellar Purkinje cell [38**,39*,40*] rather than in altered synaptic conductances is in line with the implications of the theory.

Conflict of interest statement

Nothing declared.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
 - of outstanding interest
1. Gallistel CR, Gibbon J: **Time, rate, and conditioning.** *Psychol Rev* 2000, **107**:289-344.
 2. Ward RD, Gallistel CR, Balsam PD: **It's the information!** *Behav Process* 2013, **95**:3-7.

Evidence that trials to acquisition in classical conditioning depends only on the multiplicative factor by which the expected wait time in the CS state is shorter than in the background state.

3. Gottlieb DA: **Is the number of trials a primary determinant of conditioned responding?** *J Exp Psychol Anim Behav Process* 2008, **34**:185-201.

A wildly counter-intuitive consequence of the time-scale invariance of acquisition.

4. Gibbon J, Farrell L, Locurto CM, Duncan HJ, Terrace HS: **Partial reinforcement in autoshaping with pigeons.** *Anim Learn Behav* 1980, **8**:45-59.
6. Rescorla RA, Wagner AR: **A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement.** In *Classical Conditioning II*. Edited by Black AH, Prokasy WF. Appleton-Century-Crofts; 1972:64-99.
7. Mackintosh NJ: **A theory of attention: variations in the associability of stimuli with reinforcement.** *Psychol Rev* 1975, **82**:276-298.
8. Pearce JM, Hall G: **A model for Pavlovian learning: variation in the effectiveness of conditioned but not of unconditioned stimuli.** *Psychol Rev* 1980, **87**:532-552.
9. Dragoi V, Staddon JER, Palmer RG, Buhusi CV: **Interval timing as an emergent learning property.** *Psychol Rev* 2003, **110**:126-144.
10. Killeen P: **On the temporal control of behavior.** *Psychol Rev* 1975, **82**:89-115.
11. Schultz W, Dayan P, Montague PR: **A neural substrate of prediction and reward.** *Science* 1997, **275**:1593-1599.
12. Dayan P, Kakade S, Montague PR: **Learning and selective attention.** *Nat Neurosci Suppl* 2000, **3**:1218-1223.
13. Courville AC, Daw ND, Tourtesky DS: **Bayesian theories of conditioning in a changing world.** *Trends Cogn Sci* 2006, **10**:294-300.
14. Gershman SJ, Blei DM, Niv Y: **Context, learning and extinction.** *Psychol Rev* 2010, **117**:197-209.
15. Gershman SJ, Niv JJ: **Exploring a latent cause theory of classical conditioning.** *Learn Behav* 2012, **40**:255-268.

A trial-based theory that treats conditioning as a problem in probabilistic inference.

16. Wilkes JT, Gallistel CR: **Information theory, memory, prediction, and timing in associative learning.** In *Computational Models of Brain and Behavior*. Edited by Moustafa A. Wiley/Blackwell; 2016. in press.

For a full presentation of the theory sketched in this short review

17. Rescorla RA: **Probability of shock in the presence and absence of CS in fear conditioning.** *J Comp Physiol Psychol* 1968, **66**:1-5.
 18. Gallistel CR, Craig AR, Shahan TA: **Temporal contingency.** *Behav Process* 2013, **101**:89-96.
- The application of an information-theoretic measure of contingency to classical and operant conditioning phenomena.
19. Mustaca AE, Gabelli F, Papine MR, Balsam P: **The effects of varying the interreinforcement interval on appetitive contextual conditioning.** *Anim Learn Behav* 1991, **19**:125-138.
 20. Landeira-Fernandez J, DeCola JP, Kim JJ, Fanselow MS: **Immediate shock deficit in fear conditioning: effects of shock manipulations.** *Behav Neurosci* 2006, **120**:873-879.
 21. Landeira-Fernandez J, Fanselow MS, DeCola JP, Kim JJ: **Effects of handling and context preexposure on the immediate shock deficit.** *Anim Learn Behav* 1995, **23**:335-339.

22. Bevins RA, McPhee JE, Rauhut AS, Ayres JJB: **Converging evidence for one-trial context fear conditioning with an immediate shock: importance of shock potency.** *J Exp Psychol Anim Behav Process* 1997, **23**:312-324.
23. Libby ME, Church RM: **Timing of avoidance responses by rats.** *J Exp Anal Behav* 1974, **22**:513-517.

24. Libby ME, Church RM: **Fear gradients as a function of the temporal interval between signal and aversive event in the rat.** *J Comp Physiol Psychol* 1975, **88**:911-916.
 25. Church RM, Meck WH, Gibbon J: **Application of scalar timing theory to individual trials.** *J Exp Psychol Anim Behav Process* 1994, **20**:135-155.
 26. Rakitin BC, Gibbon J, Penney TB, Malapani C, Hinton SC, Meck WH: **Scalar expectancy theory and peak-interval timing in humans.** *J Exp Psychol Anim Behav Process* 1998, **24**:15-33.
 27. Kamin LJ: **Predictability, surprise, attention, and conditioning.** In *Punishment and Aversive Behavior*. Edited by Campbell BA, Church RM. Appleton-Century-Crofts; 1969:276-296.
 28. Wagner AR, Logan FA, Haberlandt K, Price T: **Stimulus selection in animal discrimination learning.** *J Exp Psychol* 1968, **76**:171-180.
 29. Blaisdell AP, Gunther LM, Miller RR: **Recovery from blocking achieved by extinguishing the blocking CS.** *Anim Learn Behav* 1999, **27**:63-76.
 30. Matzel LD, Schachtman TR, Miller RR: **Recovery of overshadowed association achieved by extinction of the overshadowing stimulus.** *Learn Motiv* 1985, **16**:398-412.
 31. Urushihara K, Miller R: **Backward blocking in first-order conditioning.** *J Exp Psychol Anim Behav Process* 2010, **36**:281-295.
 32. Yin H, Grahame NJ, Miller RR: **Extinction of comparator stimuli during and after acquisition: differential effects on Pavlovian responding.** *Learn Motiv* 1993, **24**:219-241.
 33. Balsam PD, Drew MR, Gallistel CR: **Time and associative learning.** *Compar Cogn Behav Rev* 2010, **5**:1-22.
 34. Gallistel CR: *The Organization of Learning*. Cambridge, MA: Bradford Books/MIT Press; 1990.
 35. Gallistel CR, King AP: *Memory and the Computational Brain: Why Cognitive Science will Transform Neuroscience*. New York: Wiley/Blackwell; 2010.
 36. Wang S-H, Morris RGM: **Hippocampal-neocortical interactions in memory formation, consolidation and reconsolidation.** *Annu Rev Psychol* 2010, **61**:49-79.
 37. Gallistel CR: **The neurobiology of the language of thought.** In *Minds on Language and Thought: The Status of Cognitive Science and Its Prospects*. Edited by de Almeida RG, Gleitman L. Oxford University Press; 2016. in press.
- A radically different hypothesis about the neurobiology of memory.
38. Johansson F, Jirenhed D-A, Rasmussen A, Zucc R, Hesslow G: **Memory trace and timing mechanism localized to cerebellar Purkinje cells.** *PNAS* 2014, **111**:14930-14934.
- Strong evidence that the CS-US interval in eyeblink conditioning is encoded and remembered by molecular machinery inside cerebellar Purkinje cells.
39. Johansson F, Carlsson Hannes AE, Rasmussen A, Yeo Christopher H, Hesslow G: **Activation of a temporal memory in Purkinje cells by the mGluR7 receptor.** *Cell Rep* 2015.
- Further evidence that machinery inside cerebellar Purkinje cells encodes and retains the memory of the CS-US interval.
40. Wetmore DZ, Jirenhed D-A, Anders R, Johansson F, Schnitzer MJ, Hesslow G: **Bidirectional plasticity of Purkinje cells matches temporal features of learning.** *J Neurosci* 2014, **34**:1731-1737.
- Yet more evidence.
41. Jaynes ET: **Information theory and statistical mechanics.** *Phys Rev* 1957, **106**:620-630.
 42. Jaynes ET: *Probability Theory: The Logic of Science*. New York: Cambridge University Press; 2003.
 43. Rissanen J: **Hypothesis selection and testing by the MDL principle.** *Comput J* 1999, **42**:260-269.
 44. Grünwald PD, Myung IJ, Pitt MA: **Advances in minimum description length: theory and applications.** *Neural Information Processing Series*. Cambridge, MA: MIT Press; 2005.