Behavioural Processes 168 (2019) 103904

Contents lists available at ScienceDirect

Behavioural Processes

journal homepage: www.elsevier.com/locate/behavproc

Bayesian Behavioral Systems Theory☆

David M. Freestone[a,*], Fuat Balcı[b]

[a] William Paterson University, United States
[b] Koç University, Turkey

ABSTRACT

Behavioral Systems Theory suggests that observable behavior is embedded in a hierarchy. A CS elicits behavior because, after learning, it activates a pathway through this hierarchy. Much of Timberlake's body of work on Behavioral Systems Theory focuses on the conditions that support the conditioning of these pathways. Most notably, his work shows that the identity of the CS, US, and the CS–US interval all help support conditioning of the system. Here, we use recent experiments in the interval timing literature to motivate a Bayesian implementation of Behavioral Systems Theory. There is a probability distribution over possible pathways through the hierarchy, and the one that maximizes reinforcement is elicited. This probability distribution is conditioned on background information, like the CS–US interval and the animal's motivational state. Lower level actions of the hierarchy, like tracking prey, are conditioned on higher level goals, like the general search for food. Our implementation of Behavioral Systems Theory captures the essential features of Timberlake's verbal model; it acts as a glue, integrating sensory, timing, and decision mechanisms with observed behavior.

1. Introduction

Behavioral Systems Theory is a verbal model of animal performance in which an action is embedded in a systems hierarchy; pathways through the hierarchy that lead to a US are favored, and more easily elicited subsequently, because the US, like a reinforcement, adjusts the weights of those paths. The model is constrained by innate mechanisms, evolved over time and in response to the motivational needs of the animal: there are a fixed number of levels in the hierarchy, a fixed number of nodes at each level, and recognizably different types of reinforcement. The animal cycles through behavioral subsystems (predation, parenting, defense, etc.) based on its needs at any given time, for its entire life.

Fig. 1 shows two example paths through Timberlake's behavioral systems. Both pathways are more likely to be active when the animal is motivated to eat. In the first, the heavy gray path, the animal has not yet found food, so the pathway through 'general search' mode is more active. In the second (black), the 'focal search' mode is more active. Both pathways end with tracking prey as the observed behavior. There are many such pathways through the hierarchy, and which one becomes active depends on both motivation and reinforcement history. Many models suggest that a stimulus elicits a response, Timberlake's account is that a stimulus elicits a pathway through the hierarchy. We just observe a single response.

As an example, Timberlake and Grant (1975) conditioned a CS that predicted a food US a short time later. For one group of rats, the CS was a wooden block; for the other group, the CS was another rat. After repeated pairings, both the wooden block and the rat-CS elicited behavior more than control conditions without CS–US pairings. But the rats with the wooden block CS bit it, and the rats with the rat CS groomed it. A behavioral systems theory account is that a rat-CS conditions grooming and affiliative pathways more easily than a wooden block does. And a wooden block elicits focal search pathways more easily than a rat does. Put another way, the identity of the CS helps set what can be conditioned by reinforcement (see Garcia and Koelling, 1966; Breland and Breland, 1961).

To Timberlake, the CS supports different conditioned behaviors, adjusted by evolution. Rats will learn to press an inserted lever easily because a lever resembles, in some small way, a prey item that darts in and out of the test environment (see Timberlake, 1983). Anyone who watches a rat learn to press a lever knows that, as it learns, but before it has acquired stereotyped presses, it will sniff, jiggle, and bite the lever. During extinction, it will abandon its stereotyped pressing and sniff, jiggle, and bite the lever again. Similarly, a ball-bearing rolled into the chamber should support tracking via general search because it looks like a scurrying animal. Seeing another rat should elicit social behavior. But crucially, these behaviors are only reliably elicited by the CS after it has been paired with a US. They do not in general spontaneously occur, or at least are not sustained, without some kind of reinforcement.

A number of Timberlake's publications have focused on the

☆ We thank Mika MacInnis for introducing one of us (DMF) to Timberlake's work a decade ago, and for helpful comments and suggestions throughout the years. We also thank Peter Killeen and three anonymous reviewers for their comments and suggestions.

* Corresponding author.
E-mail address: freestoned@wpunj.edu (D.M. Freestone).

https://doi.org/10.1016/j.beproc.2019.103904
Received 23 February 2019; Received in revised form 30 June 2019; Accepted 8 July 2019
Available online 30 August 2019

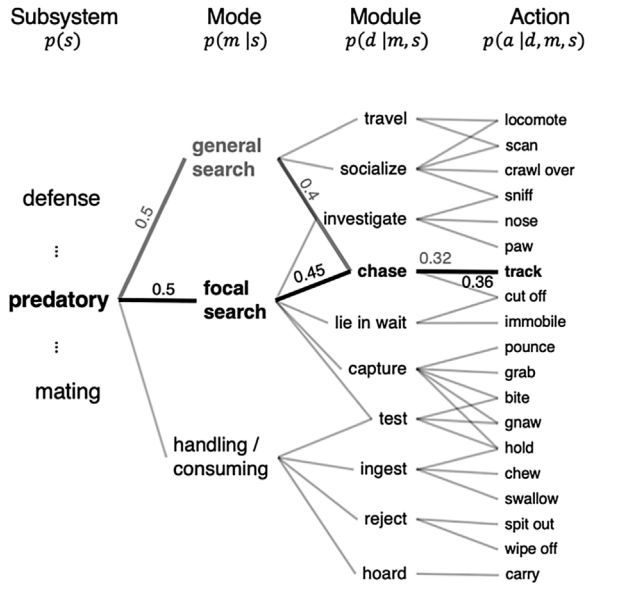

Fig. 1. In Timberlake's Behavioral Systems Theory, a single behavior elicited by a stimulus is embedded in a hierarchy. The schematic here shows two pathways through the hierarchy that lead to the same observed behavior (tracking prey). In the first, the rat tracks prey because it is in general search mode, and it tracks prey in the second case because it is in focal search mode. In our implementation of Behavioral Systems Theory, there is a probability distribution over all possible pathways, with lower levels conditioned on higher levels in the hierarchy. Background information, like the identity of the CS, US, or the CS–US interval, comes in as priors that bias the pathway that is ultimately selected. The decimals displayed in the figure show examples of how the probability of the observed action propagates through the hierarchy. In this case, the rat ultimately tracks its prey as a focal search for food. Figure adapted from Timberlake (2000).

conditions that support stimulus-elicited behavior (i.e., conditioning), and one condition in particular – the time intervals between the CSs and the US – interests us here. In one example, Timberlake et al. (1982) trained two groups of rats that a ball-bearing CS predicted a food US. The ball-bearing should support activation of the predatory subsystem (Timberlake, 1983), particularly the general and focal search modes of behavior. The CS–US interval in the two groups differed, 2 s in one group, and 6 s in the other (both relatively short intervals for the interval timing community). They found the rats with the shorter CS–US interval went directly to the food cup, and nosepoked into it. Rats with the longer CS–US interval focused on the ball bearing for a few seconds before moving to the food cup, often grabbing the ball-bearing and bringing it with them. The behavioral systems account is that short CS–US time intervals support focal search of the upcoming food, and that longer CS–US time intervals support general search for food leading to chasing and handling the prey. A second experiment found similar results (Silva and Timberlake, 1997).

In another experiment, a food pellet fell into a food cup 10-min into a 20-min session. A CS was placed immediately before it so that it co-terminated with the delivery of the pellet. After initial conditioning, six additional CSs were presented in each trial, equally spaced throughout the 20 min session (three before and three after the paired CS). These were not paired with a US. In the last phase, 19 unpaired CSs were presented, again equally spaced throughout the 20 min session (Timberlake, 2000). There were three groups of rats, each with a different CS: a ball-bearing, an inserted lever, and a light. A rat will chase a CS ball-bearing (Timberlake et al., 1982), press and gnaw at an inserted lever as a CS, and rear and approach a light CS. Timberlake and his students tracked the behavior elicited at each CS presentation.

The first result of note is that the elicited behavior to the three CSs peaked at different times throughout the 20-min session. The light-elicited responses were highest near the time of food, although the response distribution was diffuse and asymmetrical (response rates were higher before the food than after); the rats chased the ball-bearing at a high rate, the highest of all the CS-elicited behaviors, regardless of the time relative to food. Only the inserted-lever CS elicited typical timing behavior. Rats ramped up how much they contacted the lever as the time to food approached. After the food, they did not touch the lever again until the next trial. Timberlake's behavioral systems theory accounts for the result by suggesting that the conditioning that a CS supports is a joint function of the behavioral mode in the hierarchy the CS can support, and the CS–US interval.

The second result of note is that the number of CS presentations did not control the strength of the CR (and nearly any associative model suggests they must; see Gottlieb, 2008; Gottlieb and Rescorla, 2010). Contact with the light was highest when it was only presented once (phase 1), then decreased with the number of stimuli shown. But ball-bearing contacts and lever contacts did not decrease with the number of unreinforced pairings.

Timberlake's datasets are, we believe, his challenge to the timing community. Timberlake's Behavioral Systems Theory at the same time questions established beliefs about how the pairings of CSs with USs control the strength of the CR, and separately questions why the CS–US interval should shape these three behaviors in three different yet orderly and systematic ways. Behavioral Systems Theory forces us to treat different CSs that convey the exact same temporal information differently. Far from being a single stand-alone experiment, Timberlake has consistently produced experiments, stemming from Behavioral Systems Theory, that challenge the timing community in this way.

2. Timberlake's challenge to the timing community

The key challenge the timing community faces is to help answer how the CS–US relationship interacts with timing mechanisms to condition a hierarchical behavioral system. Chief among these efforts is to answer why different CS–US intervals elicit different behaviors to the same US (Silva and Timberlake, 1997), and why different CSs elicit different behaviors to the same interval (Timberlake, 2000). Further, in keeping with the spirit of both Timberlake's view and numerous experiments, the model should capture the idea that behavior is elicited by a stimulus.

Here we attempt to answer these questions. To do so, we sketch a decision-making model that takes (among other things) timing information as input and outputs an active pathway in the hierarchical behavioral system. Our model is that animals make inferences about the likelihood of various reinforcements that satisfy their motivational needs, and integrate those inferences with a value function over the possible pathways in the behavioral hierarchy. Time is inserted into the model as a mechanism that gives more plausible inferences about what reinforcements are mostly likely.

This model preserves what we believe to be the most important feature of Behavioral Systems Theory, namely, the balance between activating abstract plans like “focal search” and lower-level action commands like “track”. Reinforcement does not select responses, it biases a pathway through the systems hierarchy. Subsequent stimulus presentation do not simply call up that same pathway, it calls up the pathway that maximizes value, which may differ slightly from the reinforced one (e.g., Daw et al., 2011). In this way, it allows behavioral flexibility given the same behavioral subsystem. There is more than one way to capture prey, and more than one way to press a lever.

To motivate our implementation of Behavioral Systems Theory, we describe some results in both hierarchical reinforcement learning and temporal decision-making.

3. Hierarchical reinforcement learning

A trip to the slot machines in Vegas requires us to decide in which

casino to enter, and in which slot machine in that casino to put our money. Maybe we learn through experience that Casino A tends to pay out more than Casino B, even though, on any given occasion, the slot machine we choose in Casino A may be poor. The reinforcement earned from each pull of the slot machine should update one's value of both that slot machine and the casino, simultaneously.

Even though the reinforcement came from a single play of the slot machine, there are two prediction errors – one for the slot machine, and one for the casino. Diuk and colleagues (2013) found that participant's behavior suggested they knew both the casino's value and the value of the slot machines in that casino, and more importantly, the brain (the ventral striatum and VTA) seemed to separately encode these two predictions errors.

Standard heirarchical reinforcement learning models posit hierarchical goals. An external reinforcement like food is obtained after achieving the high level goal, and an internal pseudo-reinforcement is obtained for accomplishing a subgoal (Botvinick, 2012; Sutton et al., 1999; Dietterich, 2000). For example, a delivery truck may be paid upon delivery of the item to a house, but the driver must first pick up the package from the facility (Ribas-Fernandes et al., 2011). Picking up the package from the facility does not, in itself, lead to payment, but it is a necessary subgoal toward it. In this case, the driver knows what the subgoals are, and might use traditional reinforcement learning to learn the actions necessary to obtain the pseudo-reinforcement associated with it. To look for these pseudo-reinforcements, Ribas-Fernandez and colleagues (2011) ran an experiment based on this delivery truck analogy. As the participant maneuvered the delivery truck with a joystick to pick up the package, the package sometimes abruptly jumped to a new location, causing a pseudo-reinforcement prediction error. Ribas-Fernandez and colleagues found that the anterior cingulate cortex and habenula reflected the size of the negative prediction error (when the package jumped to a farther location), and weak evidence that the nucleus accumbens reflected the size of positive prediction error. This suggested to them that pseudo-reinforcements could be used to train hierarchical action plans. In this way, hierarchical reinforcement learning allows animals to learn both abstract high-level goals and implement them in lower-level action plans.

Because reinforcement learning uses a similar learning mechanism as the Rescorla–Wagner model – they both use the idea of a prediction error and a linear operator rule – it is tempting to use it to implement Timberlake's Behavioral Systems Theory, too. We avoid this because Timberlake's own experiments have shown that the stimulus-elicited behavior can be unrelated to the reinforcer used to condition it. A rat can be conditioned with a food reinforcer to chase a ball-bearing or groom another rat (Timberlake and Grant, 1975; Timberlake et al., 1982), but neither of these behaviors lead to the food reinforcement used to condition them. Any implementation of Behavioral Systems Theory needs a way for the properties of the stimulus and reinforcement to support conditioning of a particular path through the systems hierarchy in such a way that the elicited behaviors are not necesssarily in the service of obtaining that reinforcement.

The hierarchical reinforcement community has studied how an agent should divide a task into its goals and subgoals. In the delivery experiment, the driver knew to pick up the package from the facility because it is in the job description, but animals typically are not given explicit task instructions with the requisite subgoals explained to them. Participants could learn useful subgoals that “carve the task at its joints” (Botvinick, 2012, p. 959). To do this, they showed participants a series of images whose order of presentation was governed by an underlying transition matrix between the images that generated three clusters of images often seen near each other (Schapiro et al., 2013). A few of the items acted like critical nodes in a network that transitioned the participant between clusters. These are the environment's joints, similar to how a highway, the critical node, gives access to local neighborhoods, the clusters. Without prompting, most of the participants recognized that these nodes were special, suggesting to Shapiro and her colleagues that participants learn subgoals by learning the nodes in the structure of the task that give participants access to new clusters of states or actions in the task space. In this way, the number of levels in the hierarchy is learned through experience of the task.

In Timberlake's verbal model of Behavioral Systems, the hierarchy is fixed by evolution because the needs of the animal – predation, defense, mating, etc. – were fixed by evolution. Our implementation of the Behavioral Systems model preserves a fixed hierarchy for this reason. But we note that this is in general not the optimal hierarchy for any individual problem (Solway et al., 2014). A reasonable definition of the *optimal hierarchy* may be the hierarchy that maximizes reinforcement in the most compact way, that is, the one that minimizes the information cost required to store and use a particular hierarchy. Solway showed that the optimal hierarchy finds exactly those special nodes of the task structure described above, the ones that carve the task at its joints. Despite this, we use a fixed hierarchy for two reasons. First, we are not confident that Solway's analysis extends to classical conditioning, or to situations in which the animal has competing motivational needs that require qualitatively different reinforcement types. Second, it may be that evolution decided on a fixed hierarchy precisely because it works across many environments with fixed information cost. Nevertheless, our model below is in many ways most similar to a particular class of hierarchical reinforcement learning models (see van Dijk et al., 2011; van Dijk and Polani, 2011).

Reinforcement learning models suggest that participants choose a high level goal, and then choose sugboals that lead to it. Hierarchical reinforcement learning balances high level action plans with low level ones, and in doing so, overcomes some limitations of traditional reinforcement learning. Most notably, it helps protect against the curse of dimensionality – the fact that, as the space of possible states of the environment and possible actions is large, reinforcement learning algorithms need an unreasonably large number of training trials. For example, participants decided in what casino, and then at what slot machine, to invest. The choice of a low-level action (the slot machine) is conditioned on first choosing a casino as a high level action. Put another way, the high level choice constrains the possible actions to choose from at the levels below it; the participant cannot walk into Casino A and play one of Casino B's slot machines. In reinforcement learning models, the behavioral pathway through the hierarchy is implemented sequentially.

As far as we can tell, there is no such constraint in stimulus-elicited behavior. It is equally valid to view the pathway through the Behavioral Systems hierarchy as either (1) a series of sequential decisions that lead to a final low level action or (2) a single action defined with respect to a behavioral hierarchy. Open to either possibility, our model below is not a sequential decision-making model, but still constrains the final behavior of the animal, conditioned on the module, mode, and subsystem in the Behavioral Systems hierarchy. The second way that hierarchical reinforcement learning models protect against the curse of dimensionality is by allowing the same subgoal to be used for more than one higher goal. This is already built into Timberlake's verbal model, and thus our implementation of it below. For example, animals can track prey in either the general or focal search modes (see Fig. 1).

4. Temporal decision-making

We view timing behavior as the output of a two-stage process that first learns the time interval and then decides how to act based on it (Freestone and Church, 2016). This view has been primarily driven by a few recent papers suggesting that humans, rats, and mice optimize their reinforcement rate on temporal decision-making tasks (Balci et al., 2009, 2011; Freestone et al., 2015); and when the parameters of the task change, mice adjust their behavior quickly, often before they miss a single food pellet (Kheifets and Gallistel, 2012).

Most of these experiments use the Switch task (Balci et al., 2008), in which animals learn to switch from one option to another over time in

order to 'catch' the food pellet. On some trials, food pellets are delivered after a short interval like 3 s for the first lever press or nosepoke in one location. On the other trials, food pellets are delivered for the first press or nosepoke at the another location after a long interval, like 6 s. There is no discriminative stimulus indicating what trial type it is, so animals begin at the short location and then switch to the long location, usually after the short trial interval elapses without food. Over phases, we can manipulate the time intervals (Balci et al., 2008; Freestone et al., 2019b), reinforcement magnitudes, probabilities (Balci et al., 2009; Kheifets and Gallistel, 2012), and external variability in the food delivery times (Kheifets et al., 2017; Freestone et al., 2019a).

There is an optimal time to switch on this task, the time that balances errors against payoffs, given by Statistical Decision Theory (Wald, 1950; Blackwell and Girshick, 1979). Animals can optimize reinforcement if they choose an appropriate response distribution. On the Switch task, the switch times are normally distributed with two parameters, a mean μ and a coefficient of variation $\gamma = \sigma/\mu$. The response distribution parameterized by a normal distribution determines the types of errors (and gains) the animal is likely to make. In this case, there are four possible outcomes (Hit, Miss, False Alarm, and Correct Rejection), reducing a continuous-time Statistical Decision Theory problem to Signal Detection Theory (Green and Swets, 1966). Consider the short trial as the signal distribution, and a hit is when the animal stays long enough on the short side to catch the reinforcement on short trials. A correct rejection is when it switches in time to catch the reinforcement on long trials.

The decision the animal faces is to choose an appropriate response distribution, in this case by choosing the mean and variability of its switch times. That is, the rat seeks the solution to

$$\theta^* = \underset{\theta}{\operatorname{argmax}} \int_{\theta \in \Theta} p(r|\theta, I) v(r) d\theta$$

where r is the reward, θ is the parameter vector for the response distribution (e.g., μ and γ), $p(r|\theta, I)$ gives the probability of the reward given the response distribution and the task parameters I, and $v(r)$ gives the value of the reinforcement to the animal (for example, a simple power-law relationship may set $v(r) = kr^{\alpha}$; see Stevens, 1961). Discretizing this into the four possible outcomes gives a signal detection theory solution.

In our early work with humans and mice (and later replicated in rats), we assumed the coefficient of variation was fixed for an animal, and conditioned on this, there is an optimal (reinforcement maximizing) relationship between the mean and the coefficient of variation: animals with higher variability should have a lower mean, which is exactly what we observed (Balci et al., 2009). We showed a similar result in the Differential Reinforcement of Low Rates task (Gür et al., 2019a; Freestone et al., 2015; Çavdaroğlu et al., 2014).

When the probability of a short trial changes, mice react quickly, often before a single food pellet is missed (Kheifets and Gallistel, 2012), suggesting the mice are not using a reinforcement learning or hill-climbing strategy to adjust their behavior. If trained on the short and long trials separately with a discriminative stimulus, mice will show switch behavior on the first trial that the discriminative stimulus is removed (Tosun et al., 2016; Gür et al., 2019b), suggesting that the switch behaviour itself is not what is strengthened or reinforced. When the intervals are adjusted to make the discrimination harder (e.g., a 4 vs 12 s discrimination becomes a 4 vs 8 s discrimination), rats and mice will lower their variability, increasing their precision (Kheifets et al., 2017; Ferrara et al., 1997; Freestone et al., 2019b), suggesting the animals have some control over the variability of their response distribution and wield that control when doing so improves performance. Lastly, when external variability is added to the short and long intervals (so that they are drawn randomly each trial from a normal distribution), rats and mice on average do not adjust their switch time distribution, although many mice again decrease their variability (Kheifets et al., 2017; Freestone et al., 2019a).

Models that store in memory a single estimate of a time interval combine internal measurement error (noise in the timing mechanism) with variability in the times presented to the animal. Models with a simple transformation from these stored estimates predict more variable timing behavior. That we see, if anything, a decrease in behavioral variability suggests mice can separate internal measurement error and external variability, and that their switch response is governed in part by the internal measurement error. These findings – (i) that animals will quickly adjust their performance to maximize reinforcement rate, (ii) they can adjust their timing variability, and (iii) they separate internal and external sources of variability – suggests to us a two-stage model where a robust timing mechanism feeds into a statistical decision-making mechanism. The decision-making mechanism could be implemented in a number of different ways, including a drift-diffusion style mechanism (Simen et al., 2011; Balci and Simen, 2016, 2014).

The timing mechanism should allow for adjustable precision, separated from external variability. We suspect there are many models that fit within these constraints (including the possibility that the adjustable precision resides in the decision-mechanism), but a simple conceptual starting place is information theory. Here, the brain transmits timing information to nearby neurons via a neural code whose coding precision sets the uncertainty of the signal. Information is measured (in bits) as how precise the signal needs to be to reduce timing uncertainty to zero. This allows us to consider a probability distribution over the true time given the neural signal and its encoding precision: $p(t|\tau, \gamma)$. Rate-distortion theory (Sims, 2016; Marzen and DeDeo, 2017) or Kullback–Leibler divergences (Alexandre et al., 2019) provide a mathematically precise way for the brain to set and adjust its timing precision based on the task.

Another reason to prefer a probabilistic treatment of the timing mechanism is that it provides the framework for a probabilistic behavioral systems hierarchy that feeds into statistical decision theory.

4.1. Timescale invariance

Weber's law for interval timing (Gibbon, 1977), is that the standard deviation with which we respond at a time interval scales linearly with that interval such that the coefficient of variation is constant, $\sigma/\mu = \gamma$. A stronger form of the scalar property, timescale invariance, is often visible when the time axis for these behaviors is normalized by the interval (e.g., divide all the timestamped events of a 10 s CS–US interval by 10, and all the timestamped events of a 100 s CS–US interval by 100 s), the response distributions completely overlap, suggesting that the same pattern of behaviors are simply stretched out in time based on the CS–US interval.

In one experiment, Silva and Timberlake (1998) trained rats on a long interval filled with clock CSs (light CSs that flashed faster with each presentation approaching the food). They measured behaviors, classified as general search, focal search, postfood focal search, and food consumption/handling, and found that general and focal search behavior scaled with the interval, but that postfood behaviors were in their words time bound; they always occurred at roughly the same times since the previous food delivery.

From this, Silva and Timberlake argue that the behavioral modes are timescale invariant, all except food handling/consumption, which takes a fixed amount of absolute time (Others have shown that it may be related to meal size and the time it takes to prepare it; Krebs et al., 1977). As far as we know, there is no evidence in favor of or against timescale invariance in subsystems other than predatory, like defense or mating.

If true, it suggests that the brain's interval timing mechanism may feed into the behavioral systems hierarchy to support conditioning. Models of interval timing account for timescale invariance in different ways: the clock rate is set by motivation and the reinforcement rate (Killeen and Fetterman, 1988; Machado, 1997), storing intervals in memory is noisy (Gibbon et al., 1984), decision processes are noisy

(Guilhardi et al., 2007), or that it comes from balanced excitation and inhibition in neural populations (Simen et al., 2011). The information theoretic view we take here can produce timescale invariant behavior by using a floating-point coding scheme (Gallistel, 2018).

5. Hierarchical behavioral systems

Behavioral systems theory posits a fixed number of levels in a behavioral hierarchy, and a fixed number of nodes within each level. This gives the animal a flexible but limited repertoire of behaviors. Our implementation of behavioral systems theory further assumes that animals want to maximize their reinforcement, although it would seem natural to extend this to maximizing total *future* reinforcement (Sutton and Barto, 1998). Reinforcement can come from many sources, each of which satisfy a basic need for the animal, e.g., eating, drinking, mating, socializing, etc. These basic needs form Timberlake's top level *subsystem* of the behavioral systems hierarchy. If all reinforcements were worth the same, maximizing reinforcement in the long run would reduce to choosing the pathway through the hierarchy that is most consistent with reinforcement.

Let the state of the rat in its environment be described by the joint distribution $p(r, \theta, I)$, where r is a reinforcement (food, sex, social reinforcement, etc.), θ is a pathway through the hierarchy, and I is task relevant information obtained through learning or the sensory system, information like identity of the CS, US, and the CS–US interval. This probability can be decomposed into a set of conditional probabilities

$$p(r, \theta, I) = p(r|\theta, I)p(\theta|I)p(I)$$

the probability of reinforcement r given the pathway and background information, the probability of the pathway given background information, and the prior probability of background information. The parameter vector θ gives the levels and nodes in the systems hierarchy, the subsystem (s), mode (m), module (d) and the action (a): $\theta = \{s, m, d, a\}$. This joint probability could be partitioned into a set of conditional probabilities a number of ways. The probability of having two brown eyes is the same regardless of which eye we check first. But the order is crucial for us here. There is little point in performing a focal search for food if the animal is in a mating subsystem. The joint distribution over θ is decomposed into sets of conditional probabilities where the lower levels of the hierarchy are conditioned on the levels above it

$$p(\theta) = p(a|d, m, s)p(d|m, s)p(m|s)p(s)$$

It is exactly this feature that allows our implementation of Behavioral systems theory to balance a motivational goal with a lower-level action: it has the potential to elicit in our rat many different actions consistent with the module, many different modules consistent with the mode, many different modes consistent with the subsystem, a subsystem consistent with the animal's motivational state M, and all consistent with background information I from its environment.

Timing information is injected into the hierarchy through background information I. We suggest that the animal computes the probability distribution over the possible reinforcement times in its environment, contingent on the time markers in its environment, e.g., CS onset or termination. For example, if a rat is motivated to eat, $p(M = \text{eat})$ is high. If food is nearby, $p(\text{food}|\tau, \gamma)$ is also high. Together, these bias the systems hierarchy (θ) toward the predatory system. In this way, a pathway can still be active at the wrong time, provided there is enough motivation for it: A very hungry rat may still search for food even if it has learned food is unlikely to be close.

Motivation also biases time perception: $p(I) \propto p(t|\tau, \gamma, M)p(M)$. This allows motivation to adjust an animal's precision (Kheifets et al., 2017; Freestone et al., 2019b), or its interval estimate. At present, we suppose that changes to the mean of an animal's timed behavior – for example earlier start times on a Peak procedure or similar (Balci et al., 2010a,b; Ludvig et al., 2011; Balci, 2014; Fox and Kyonka, 2014; Galtress et al., 2012) – are caused not by shifts in time estimation but by the changes that $p(M)$ has on the joint distribution $p(r, \theta, I)$ or the value function (see Freestone and Church, 2010; Taylor et al., 2007).

While the value function may be of any form, a reasonable starting place may be Steven's power law (Stevens, 1961): $v(r) = \text{kr}^{\alpha}$, where α may further be some function of motivation, energy constraints, etc. The pathway through the hierarchy – leading to the action ultimately observed – is the pathway that maximizes reinforcement for the animal, θ^*, computed using statistical decision theory, although Rate-distortion (Sims, 2016; Marzen and DeDeo, 2017) or the Kullback–Leibler divergence (Alexandre et al., 2019) provide similar ways of integrating information theoretic probability distributions with value functions.

Reinforcement is used in the Reinforcement Learning sense and is different from a US. A reinforcement is an event that has value to the animal and fulfills one of its basic needs. In this way, a food US is both a stimulus conveying information to the animal (and comes in as I), and a food reinforcement (and comes in as r). The animal seeks to maximize reinforcement, not US presentations. This is needed to explain why a rat will groom a rat-CS more after it has been paired with a food-US. Using statistical decision theory, the value of a particular pathway is computed across all reinforcements, that is, by integrating them out

$$v(\theta, I) = \sum_{r \in R} p(r, \theta, I)v(r)$$

Learning the pathways that lead to reinforcement is accomplished by solving the inverse inference problem, namely, by updating $p(\theta, I|r)$ which can be computed using Bayes' theorem.

5.1. Explanation of results

The three findings we seek to explain with our model are: (1) elicited behavior sometimes matches the CS (e.g., rats groom a rat-CS even if the US is food; Timberlake and Grant, 1975) and sometimes matches the US (the way pigeons peck a key is related to the upcoming US; Jenkins and Moore, 1973); (2) the same stimulus can elicit different behaviors depending on the CS–US interval (Silva and Timberlake, 1997; Timberlake, 2000); and (3) the response rate or elicitation probability does not (always) depend on the number of CS–US pairings (Gottlieb, 2008; Gottlieb and Rescorla, 2010; Timberlake, 2000).

Timberlake's accounts of elicited behavior usually appeals to how external factors bias the behavioral system. Our model preserves that feature by claiming that external factors enter into the model through background information I. This information sets the conditions for biasing the pathways through the behavioral hierarchy, $p(\theta|I)$. In this way, our implementation of Behavioral Systems Theory is not a model of reinforcement learning, CS identification, or timing – even though our choice of model was informed by the timing literature described above.

An important feature of our model is that it attempts to describe stimulus elicited behavior, not the value of actions. The active pathways are described by the joint function $p(r, \theta, I)$ rather than the first term in its decomposition, $p(r|\theta, I)$. When a CS turns on, information about the CS and the US come in through the background information I. With a lever-CS, treated like a small prey item to a rat, the two likely reinforcements are both food: the lever-as-food and the upcoming food-US it predicts. After integrating over the possible reinforcements, a hierarchical pathway through the predatory system directed at the lever is more valuable. With a rat-CS, the background information is that the CS is a rat and that its presence may reduce the time to a social reinforcement. The two likely reinforcements are thus social and food. After integrating over the possible reinforcements, a hierarchical pathway through an affiliative system directed toward the rat-CS is more valuable. This is another reason why our model is not a simple rehash of already successful hierarchical reinforcement learning models mentioned above. Behavior is elicited by a CS, and that behavior need not be directed toward the US. Reinforcements invigorate each other (e.g., Sheffield and Campbell, 1954; Niv et al., 2007).

The second of Timberlake's results – the same stimulus can elicit different behaviors depending on their relative time to food – is explained again with respect to learning the background information, I. At the time of CS onset, which is all we are interested in here, the animal computes the probability distribution over possible reinforcement times given its encoding τ, precision γ, and reinforcements $r \in R$. Far from the food US, the probability of a food reinforcement is low, so the active pathways favor general search by chasing the ball-bearing. Close to food, the probability of food reinforcement is high, so the active pathway shifts to focal search in the food cup.

Timing information can be injected into the systems hierarchy by assuming an information theoretic code for communicating timing information, which naturally gives rise to a probability distribution for time that can bias pathways through the behavioral system. Other timing models are in principle consistent with the Behavioral Systems model we describe here, provided they output not a response or an association, but a way of biasing the active pathway through the hierarchy. This bias need not be based on probabilities, but it is the most natural input to our model.

The third of Timberlake's results – the response rate or elicitation probability does not (always) depend on the number of CS–US pairings – is easiest to explain. Our model is not an associative model (e.g., Rescorla and Wagner, 1972), nor does it rely on the cached value of actions (e.g., Sutton and Barto, 1998). We do not use reinforcement to increase or decrease weights between nodes in a connectionist network. Instead, we draw on probabilistic inference and information theory, which suggest that the number of training trials do not influence the response rate (Ward et al., 2013; Balsam and Gallistel, 2009; Gallistel and Balsam, 2014).

In general, formulating Behavioral Systems Theory as probabilistic inference allows us to ask more general questions about the modes (or modules, or subsystem) that the CS elicits. Given the joint probability distribution $p(r, \theta, I)$, the probability that a stimulus elicits behavioral mode m is just the marginal distribution for m, conditioned on all the levels above it, and averaged over all the layers beneath.

Our implementation of Behavioral Systems Theory makes no predictions or claims about response rate. Most models do not. Scalar expectancy theory is silent on the issue altogether, except to output whether the animal is in a low or high response state (Gibbon et al., 1984; Church et al., 1994). Most associative models also do not directly model response rate either, and instead suggest that response rate is some linear (or at least monotonic) function of reinforcement rate (Rescorla and Wagner, 1972; Killeen and Fetterman, 1988; Machado, 1997). Similarly, a modular theory of learning and performance (Guilhardi et al., 2007) outputs a *response tendency* metric that feeds into a response module that emits packets of a few responses at a time, each with Wald-distributed interresponse times. The parameters of their Wald distribution were obtained by fitting previous experiments, rather than on theoretical grounds. Niv's response vigor model suggests that animals make two decisions, first what action to make and second how long to maintain the action (Niv et al., 2006b,a, 2007). As soon as that action is finished, the animal decides on another action, and again for how long. It could decide to repeat the action (or persist at it for a longer duration), which increases response rate. The reinforcement rate, and motivation in general, helps set how long an action lasts. We see the appeal of both Modular Theory's packet of responses and Niv's 'how long' decision.

5.2. Psychological and neurobiological plausibility

The plausibility of this model rests on a brain's ability to assign probabilities and conditional probabilities to events, and to transmit information via a compressed neural code. This hypothesis has been heavily studied in psychology and neuroscience for over half a century (Attneave, 1954; Barlow, 1961; Miller, 1956; Hick, 1952), and forms a core thread in the neuroscience community as a unified theory of both brain structure and function (Doya et al., 2011; Rieke et al., 1999; Sterling and Laughlin, 2017). Although not without its critics, this view has lived up to scrutiny and analysis. Brains can in principle implement Bayes theorem (Darlington et al., 2018; Knill and Pouget, 2004; Ma et al., 2006; Pouget et al., 2013) and a spike train can be viewed transmitting a compressed (and lossy) information theoretic code (Qian and Zhang, 2019). Area V1 seems to adjust its orientation tuning curves to fit the prior distribution of angles in its environment (Girshick et al., 2011; Goris et al., 2015). And in timing tasks, animals take into account the prior probability of intervals they are likely to see (Jazayeri and Shadlen, 2010; Shi et al., 2013). With the exception of a floating-point representation for timing information (Gallistel, 2018) – which, as far as we know, does not always implement an efficient code – we believe that very little in our model would be controversial to this large community (although it may be controversial to some outside it).

5.3. Fitting the model to data

Timberlake's verbal model of Behavioral Systems Theory acts as a module whose input is external information about the CS and the US – including their identities and when they occur in time – and whose output is an action defined with respect to a systems hierarchy. In this way, Timberlake's model acts as a glue that connects sensory mechanisms to stimulus-elicited behavior. This means that our implementation of the model should in principle be fit to data, provided a specification describing how sensory information is translated into probability, and provided the right kind of data are collected. We have spent considerable space in this paper describing why we believe that background information should enter as prior information into a Bayesian hierarchical system.

Collecting the right kind of data is a harder problem. In most classical conditioning experiments, data about a single CR is collected, e.g., a head-entry or key-peck. And in most operant chambers, this is all that *can* be collected. Timberlake's lab has collected data primarily using food USs that condition behaviors with up to three CSs and a few time intervals. Because our model uses hierarchical probabilities, it is in principle possible to fit the parts of the model for which we do have data, and drop the parts of the model for which we do not. For example, Timberlake assumes in most of his experiments that the animal is in a predatory subsystem when they enter the chamber, removing the need to fit parameters in any other subsystem. When only a few types of behaviors are measured, we can remove the others from the action level of the hierarchy, and the probabilities renormalize to fill the space of behaviors we did measure. Modern fitting tools like the Stan programming language (Carpenter et al., 2017) make short work of hierarchical models with many parameters, including hidden Markov models of the form we described here.

6. Conclusion

Here we described one possible implementation of Timberlake's Behavioral Systems Theory. We took special care to stay as close to Timberlake's verbal model as possible because we see considerable value in it. Otherwise, our model was primarily informed by our recent work on interval timing and decision-making mechanisms in both rats and mice. This pushed us to consider a model in which an information theoretic code transmits timing information throughout the brain, which can be described by a probability distribution for time intervals that, because of the code used, obeys Weber's Law. Our previous work suggested to us that the probability distribution for time intervals is then fed into statistical decision theory that leads to an animal's response distribution. To model a systems hierarchy, this same general apparatus – statistical decision theory that operates on a distribution over possible responses – was simply expanded to a distribution over a hierarchical behavioral system.

The features we hoped to retain from Timberlake's verbal model

were (1) it balances high level goals with low level actions, giving the animal the flexibility to choose many behaviors consistent with a single goal; (2) that information about the CS, US, and the time between them supports the conditioning of particular behaviors over others; and (3) modularity: even though we try to give specifics about how background information could bias the systems hierarchy, we explicitly noted in several places where other mechanisms may do the job just as well. That is, we tried to retain Timberlake's openness toward models, allowing for a modular approach where many models could be tested without any change to our implementation to Behavioral Systems Theory.

References

Alexandre, Z., Oleg, S., Giovanni, P., 2019. An information-theoretic perspective on the costs of cognition. Neuropsychologia 123, 5–18.

Attneave, F., 1954. Some informational aspects of visual perception. Psychol. Rev. 61 (3), 183–193. https://doi.org/10.1037/h0054663.

Balci, F., 2014. Interval timing, dopamine, and motivation. Timing Time Percept. 2 (3), 379–410.

Balci, F., Freestone, D., Gallistel, C.R., 2009. Risk assessment in man and mouse. Proc. Natl. Acad. Sci. U.S.A. 106 (7), 2459–2463. https://doi.org/10.1073/pnas.0812709106.

Balci, F., Freestone, D., Simen, P., deSouza, L., Cohen, J.D., Holmes, P., 2011. Optimal temporal risk assessment. Front. Integr. Neurosci. 5. https://doi.org/10.3389/fnint.2011.00056.

Balci, F., Ludvig, E.A., Abner, R., Zhuang, X., Poon, P., Brunner, D., 2010a. Motivational effects on interval timing in dopamine transporter (DAT) knockdown mice. Brain Res. 1325, 89–99.

Balci, F., Ludvig, E.A., Brunner, D., 2010b. Within-session modulation of timed anticipatory responding: when to start responding. Behav. Process. 85 (2), 2204–2206.

Balci, F., Papachristos, E.B., Gallistel, C.R., Brunner, D., Gibson, J., Shumyatsky, G.P., 2008. Interval timing in genetically modified mice: a simple paradigm. Genes Brain Behav. 7 (3), 2373–2384. https://doi.org/10.1111/j.1601-183X.2007.00348.x.

Balci, F., Simen, P., 2014 June. Decision processes in temporal discrimination. Acta Psychol. 149, 157–168. https://doi.org/10.1016/j.actpsy.2014.03.005.

Balci, F., Simen, P., 2016. A decision model of timing. Curr. Opin. Behav. Sci. 8, 94–101.

Balsam, P.D., Gallistel, C.R., 2009. Temporal maps and informativeness in associative learning. Trends Neurosci. 32 (2), 273–278.

Barlow, H.B., 1961. Possible principles underlying the transformation of sensory messages. Sensory Commun. 1, 217–234.

Blackwell, D.A., Girshick, M.A., 1979. Theory of Games and Statistical Decisions. Courier Corporation.

Botvinick, M.M., 2012. Hierarchical reinforcement learning and decision making. Curr. Opin. Neurobiol. 22 (6), 2956–2962.

Breland, K., Breland, M., 1961. The misbehavior of organisms. Am. Psychol. 16 (11), 2681–2684. https://doi.org/10.1037/h0040090.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Riddel, A., 2017. Stan: a probabilistic programming language. J. Stat. Softw. 76 (1), 21–32.

Çavdaroğlu, B., Zeki, M., Balci, F., 2014. Time-based reward maximization. Philos. Trans. R. Soc. B: Biol. Sci. 369, 1637. https://doi.org/10.1098/rstb.2012.0461.

Church, R.M., Meck, W.H., Gibbon, J., 1994. Application of scalar timing theory to individual trials. J. Exp. Psychol.: Anim. Behav. Process. 20 (2), 2135.

Darlington, T.R., Beck, J.M., Lisberger, S.G., 2018. Neural implementation of Bayesian inference in a sensorimotor behavior. Nat. Neurosci. 21 (10), 21442. https://doi.org/10.1038/s41593-018-0233-y.

Daw, N.D., Gershman, S.J., Seymour, B., Dayan, P., Dolan, R.J., 2011. Model-based influences on humans' choices and striatal prediction errors. Neuron 69 (6), 21204–21215. https://doi.org/10.1016/j.neuron.2011.02.027.

Dietterich, T.G., 2000. Hierarchical reinforcement learning with the MAXQ value function decomposition. J. Artif. Intell. Res. 13, 227–303.

Diuk, C., Tsai, K., Wallis, J., Botvinick, M., Niv, Y., 2013. Hierarchical learning induces two simultaneous, but separable, prediction errors in human basal ganglia. J. Neurosci. 33 (13), 25797–25805. https://doi.org/10.1523/JNEUROSCI.5445-12.2013.

Doya, K., Ishii, S., Pouget, A., Rao, R.P.N. (Eds.), 2011. Bayesian Brain: Probabilistic Approaches to Neural Coding. The MIT Press, Cambridge, MA.

Ferrara, A., Lejeune, H., Wearden, J.H., 1997. Changing sensitivity to duration in human scalar timing: an experiment, a review, and some possible explanations. Quart. J. Exp. Psychol. B: Comp. Physiol. Psychol. 50B (3), 2217–2237.

Fox, A.E., Kyonka, E.G., 2014. Choice and timing in pigeons under differing levels of food deprivation. Behav. Process. 106, 82–90.

Freestone, D., Antonellis, D., Kennedy, T., Bzadough, N., 2019a. Mice separate internal measurement error and external variability (submitted for publication).

Freestone, D.M., Balci, F., Simen, P., Church, R.M., 2015. Optimal response rates in humans and rats. J. Exp. Psychol. Anim. Learn. Cogn. 41 (1), 239–251. https://doi.org/10.1037/xan0000049.

Freestone, D.M., Church, R.M., 2010 May. The importance of the reinforcer as a time marker. Behav. Process. 84 (1), 2500–2505. https://doi.org/10.1016/j.beproc.2010.01.011.

Freestone, D.M., Church, R.M., 2016. Optimal timing. Curr. Opin. Behav. Sci. 8, 276–281. https://doi.org/10.1016/j.cobeha.2016.02.031.

Freestone, D.M., Donskoy, B., Sari, D., Rosenberg, C., 2019b. Temporal measurement error is sensitive to task difficulty (submitted for publication).

Gallistel, C.R., Balsam, P.D., 2014. Time to rethink the neural mechanisms of learning and memory. Neurobiol. Learn. Memory 108, 136–144. https://doi.org/10.1016/j.nlm.2013.11.019.

Gallistel, C.R., 2018. Finding numbers in the brain. Philos. Trans. R. Soc. B: Biol. Sci. 373 (1740), 220170119. https://doi.org/10.1098/rstb.2017.0119.

Galtress, T., Marshall, A.T., Kirkpatrick, K., 2012. Motivation and timing: clues for modeling the reward system. Behav. Process. 90 (1), 2142–2153.

Garcia, J., Koelling, R.A., 1966. Relation of cue to consequence in avoidance learning. Psychon. Sci. 4 (1), 2123–2124. https://doi.org/10.3758/BF03342209.

Gibbon, J., 1977. Scalar expectancy theory and Weber's law in animal timing. Psychol. Rev. 84 (3), 2279–2325. https://doi.org/10.1037/0033-295X.84.3.279.

Gibbon, J., Church, R.M., Meck, W.H., 1984. Scalar timing in memory. Ann. N. Y. Acad. Sci. 423 (1), 252–277.

Girshick, A.R., Landy, M.S., Simoncelli, E.P., 2011 July. Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. Nat. Neurosci. 14 (7), 2926–2932. https://doi.org/10.1038/nn.2831.

Goris, R.L., Simoncelli, E.P., Movshon, J.A., 2015. Origin and function of tuning diversity in macaque visual cortex. Neuron 88 (4), 2819–2831. https://doi.org/10.1016/j.neuron.2015.10.009.

Gottlieb, D.A., 2008. Is the number of trials a primary determinant of conditioned responding? J. Exp. Psychol. Anim. Behav. Process. 34 (2), 2185–2201. https://doi.org/10.1037/0097-7403.34.2.185.

Gottlieb, D.A., Rescorla, R.A., 2010. Within-subject effects of number of trials in rat conditioning procedures. J. Exp. Psychol. Anim. Behav. Process. 36 (2), 2217–2231. https://doi.org/10.1037/a0016425.

Green, D.M., Swets, J.A., 1966. Signal Detection Theory and Psychophysics. Wiley, New York (OCLC: 890266).

Guilhardi, P., Yi, L., Church, R.M., 2007. A modular theory of learning and performance. Psychon. Bull. Rev. 14 (4), 2543–2559.

Gür, E., Fertan, E., Kosel, F., Wong, A.A., Balci, F., Brown, R.E., 2019a. Sex differences in the timing behavior performance of 3xtg-AD and wild-type mice in the peak interval procedure. Behav. Brain Res. 360, 235–243.

Gür, E., Duyan, Y.A., Balcı, F., 2019b. Probabilistic information modulates the timed response inhibition deficit in aging mice. Front. Behav. Neurosci. https://doi.org/10.3389/fnbeh.2019.00196.

Hick, W.E., 1952. On the rate of gain of information. Quart. J. Exp. Psychol. 4 (1), 211–226. https://doi.org/10.1080/17470215208416600.

Jazayeri, M., Shadlen, M.N., 2010. Temporal context calibrates interval timing. Nat. Neurosci. 13 (8), 21020–21026. https://doi.org/10.1038/nn.2590.

Jenkins, H.M., Moore, B.R., 1973. The form of the auto-shaped response with food or water reinforcers. J. Exp. Anal. Behav. 20 (2), 2163–2181. https://doi.org/10.1901/jeab.1973.20-163.

Kheifets, A., Freestone, D., Gallistel, C.R., 2017 July. Theoretical implications of quantitative properties of interval timing and probability estimation in mouse and rat. J. Exp. Anal. Behav. 108 (1), 239–272. https://doi.org/10.1002/jeab.261.

Kheifets, A., Gallistel, C.R., 2012. Mice take calculated risks. Proc. Natl. Acad. Sci. U.S.A. 109 (22), 28776–28779.

Killeen, P.R., Fetterman, J.G., 1988. A behavioral theory of timing. Psychol. Rev. 95 (2), 2274.

Knill, D.C., Pouget, A., 2004. The Bayesian brain: the role of uncertainty in neural coding and computation. Trends Neurosci. 27 (12), 2712–2719. https://doi.org/10.1016/j.tins.2004.10.007.

Krebs, J.R., Erichsen, J.T., Webber, M.I., Charnov, E.L., 1977. Optimal prey selection in the great tit (*Parus major*). Anim. Behav. 25, 30–38.

Ludvig, E.A., Balci, F., Spetch, M.L., 2011. Reward magnitude and timing in pigeons. Behav. Process. 86 (3), 2359–2363.

Ma, W.J., Beck, J.M., Latham, P.E., Pouget, A., 2006. Bayesian inference with probabilistic population codes. Nat. Neurosci. 9 (11), 21432.

Machado, A., 1997. Learning the temporal dynamics of behavior. Psychol. Rev. 104 (2), 2241.

Marzen, S.E., DeDeo, S., 2017. The evolution of lossy compression. J. R. Soc. Interface 14 (130), 220170166.

Miller, G.A., 1956. The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol. Rev. 63 (2), 281–297. https://doi.org/10.1037/h0043158.

Niv, Y., Daw, N.D., Dayan, P., 2006a. How fast to work: response vigor, motivation and tonic dopamine. Adv. Neural Inform. Process. Syst. 1019–1026.

Niv, Y., Daw, N.D., Joel, D., Dayan, P., 2007. Tonic dopamine: opportunity costs and the control of response vigor. Psychopharmacology 191 (3), 2507–2520.

Niv, Y., Joel, D., Dayan, P., 2006b. A normative perspective on motivation. Trends Cogn. Sci. 10 (8), 2375–2381.

Pouget, A., Beck, J.M., Ma, W.J., Latham, P.E., 2013. Probabilistic brains: knowns and unknowns. Nat. Neurosci. 16 (9), 21170–21178. https://doi.org/10.1038/nn.3495.

Qian, N., Zhang, J., 2019. Neuronal firing rate as code length: a hypothesis. Comput. Brain Behav. https://doi.org/10.1007/s42113-019-00028-z.

Rescorla, R.A., Wagner, A.R., 1972. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. Class. Condition. II: Curr. Res. Theory 2, 64–99.

Ribas-Fernandes, J.J.F., Solway, A., Diuk, C., McGuire, J.T., Barto, A.G., Niv, Y., Botvinick, M.M., 2011 July. A neural signature of hierarchical reinforcement learning. Neuron 71 (2), 2370–2379. https://doi.org/10.1016/j.neuron.2011.05.042.

Rieke, F., Warland, D., Steveninck, R.d.R.v., Bialek, W., 1999. Spikes: Exploring the

Neural Code (Reprint Edition). A Bradford Book, Cambridge, MA.
Schapiro, A.C., Rogers, T.T., Cordova, N.I., Turk-Browne, N.B., Botvinick, M.M., 2013. Neural representations of events arise from temporal community structure. Nat. Neurosci. 16 (4), 2486–2492. https://doi.org/10.1038/nn.3331.
Sheffield, F.D., Campbell, B.A., 1954. The role of experience in the spontaneous activity of hungry rats. J. Comp. Physiol. Psychol. 47 (2), 100–297.
Shi, Z., Church, R.M., Meck, W.H., 2013. Bayesian optimization of time perception. Trends Cogn. Sci. 17 (11), 2556–2564. https://doi.org/10.1016/j.tics.2013.09.009.
Silva, K.M., Timberlake, W., 1997. Behavior systems view of conditioned states during long and short CS–US intervals. Learn. Motiv. 28 (4), 2465–2490. https://doi.org/10.1006/lmot.1997.0986.
Silva, K.M., Timberlake, W., 1998. The organization and temporal properties of appetitive behavior in rats. Anim. Learn. Behav. 26 (2), 2182–2195.
Simen, P., Balci, F., Cohen, J.D., Holmes, P., 2011. A model of interval timing by neural integration. J. Neurosci. 31 (25), 29238–29253.
Sims, C.R., 2016. Rate-distortion theory and human perception. Cognition 152, 181–198.
Solway, A., Diuk, C., Córdova, N., Yee, D., Barto, A.G., Niv, Y., Botvinick, M.M., 2014. Optimal behavioral hierarchy. PLOS Comput. Biol. 10 (8), 2e1003779. https://doi.org/10.1371/journal.pcbi.1003779.
Sterling, P., Laughlin, S., 2017. Principles of Neural Design (Reprint Edition). The MIT Press.
Stevens, S.S., 1961. To honor Fechner and repeal his law: a power function, not a log function, describes the operating characteristic of a sensory system. Science 133 (3446), 280–286. https://doi.org/10.1126/science.133.3446.80.
Sutton, R.S., Barto, A.G. (Eds.), 1998. Reinforcement Learning: An Introduction. MIT Press, Cambridge, MA, USA.
Sutton, R.S., Precup, D., Singh, S., 1999. Between MDPs and semi-MDPs: a framework for temporal abstraction in reinforcement learning. Artif. Intell. 112 (1), 2181–2211. https://doi.org/10.1016/S0004-3702(99)00052-1.
Taylor, K.M., Horvitz, J.C., Balsam, P.D., 2007. Amphetamine affects the start of responding in the peak interval timing task. Behav. Process. 74 (2), 2168–2175. https://doi.org/10.1016/j.beproc.2006.11.005.
Timberlake, W., 1983. Rats' responses to a moving object related to food or water: a behavior-systems analysis. Anim. Learn. Behav. 11 (3), 2309–2320. https://doi.org/10.3758/BF03199781.
Timberlake, W., 2000. Motivational modes in behavior systems. Handbook of Contemporary Learning Theories. Psychology Press, pp. 165–220.
Timberlake, W., Grant, D.L., 1975. Auto-shaping in rats to the presentation of another rat predicting food. Science 190 (4215), 2690–2692.
Timberlake, W., Wahl, G., King, D., 1982. Stimulus and response contingencies in the misbehavior of rats. J. Exp. Psychol. Anim. Behav. Process. 8 (1), 262–285.
Tosun, T., Gur, E., Balci, F., 2016. Mice plan decision strategies based on previously learned time intervals, locations, and probabilities. Proc. Natl. Acad. Sci. U.S.A. 113 (3), 2787–2792.
van Dijk, S.G., Polani, D., 2011. Grounding subgoals in information transitions. 2011 IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning (ADPRL) 105–111. https://doi.org/10.1109/ADPRL.2011.5967384.
van Dijk, S.G., Polani, D., Nehaniv, C.L., 2011. Hierarchical behaviours: getting the most bang for your bit. In: Kampis, G., Karsai, I., Szathmáry, E. (Eds.), Advances in Artificial Life. Darwin Meets von Neumann. Springer Berlin Heidelberg, pp. 342–349.
Wald, A., 1950. Statistical Decision Functions. Wiley, Oxford, England.
Ward, R.D., Gallistel, C.R., Balsam, P.D., 2013 May. It's the information!. Behav. Process. 95, 3–7. https://doi.org/10.1016/j.beproc.2013.01.005.