The Future of the Behavioral Sciences

Arturo Bouzas, Carlos Velázquez, Manuel Villareal

Every science requires a manifest, and psychology is no exception. In 1913, James Watson wrote the following: "*Psychology [...] is a purely objective experimental branch of natural science. Its theoretical goal is the prediction and control of behavior. It recognizes no dividing line between man and brute. The behavior of man, with all of its refinement and complexity, forms only a part of the behaviorist's total scheme of investigation.*" We omitted from the quote "*psychology as seen by a behaviorists*" to clarify that psychology as a science is defined by its commitment to the practices defining any other natural science, and it has nothing to do with a particular "school."

Since its beginnings as an experimental science in the 19th century, that studied the response to physical stimuli, psychology has had a close relationship to mathematics that is well summarized in the three volumes of Mathematical Psychology published in 1963 (R. D. Luce et al.), in recent review papers (Falmagne & Lee, 2002; D. Luce, 1995) and in two new handbooks (Batchelder, Colonius, Dzhafarov, & Myung, 2016). It is in this context that we are so honored to participate in a tribute to José Antonio de la Peña. No mexican mathematician has understood and better appreciated the relation between mathematics and psychology than he did. His broad interests led him to forage in areas such as game theory and to recognize that mathematics, as a language of nature, also covers the study of the behavior of living beings.

The present chapter is organized in the following way. In the first part we will argue in favor of studying behavior as an adaptation to the statistical structure of the environment. In the second part we will present an example of how a very simple mathematical model of an integrator can describe a very general principle of adaptation common to bacteria and human beings. In the third part we will extend the model to an environmental structure characterized by changes with noise. We will conclude with a brief discussion about the consequences of new developments in statistical techniques in the formation of behavioral scientists.

Throughout its history, psychology has argued about the "natural class" that defines its object of study. We will argue for a naturalistic answer. The science of behavior studies the adapted behavior as well as the mechanisms that have an origin in natural selection and that generate adaptive behavior. Biological organisms expend energy and require a constant resupply of it. Evolutionary forces work as a filter that shape and select successful forms of obtaining energy constrained by the distribution of the sources (biologically important events or BIE). These elements

conform an empirical system that is the proper object of study of the science of behavior: an adaptive and adapted system. These systems can be seen as possible optimal solutions to problems associated with higher reproductive success. It is important to consider that there is no adapting agent; we are instead talking about a process without agency or intention, a process of selection under constraints.

To make sense of the notion of adaptability as an optimization process we need to start by identifying the set of constraints under which the resupply of energy occurs. There can be two types of constraints. The first is associated with the organism and the second with the statistical properties of the BIE. Among the first, the most important is that all behavior expends energy and takes time: therefore, different behaviors compete for the available time and energy. Other important organismic constraints are the historical (possible solutions acquired during the lifespan of the organism), computational and physiological (mechanisms that allow detecting, discriminating and acting on BIE).

The second set of constraints are those associated with the distribution of the BIE in the environment, what we call the statistical properties of the environment. The availability of the BIE can be associated with the following: 1) time, 2) location, 3) covariance with other environmental events, and 4) covariance with the behavior of the organism. Organisms that can detect and learn about these properties will be in a better position to distribute their behavior optimally to obtain the different relevant BIE. These constraints have been the object of much research in timing, place learning, classical conditioning and instrumental conditioning and have been an important source of psychological knowledge (Bouton, 2016; Staddon, 2016).

The previous constraints are limited to individual events, discrete stimuli and responses. Nevertheless, organisms can detect and adapt to second order statistical properties of the distribution of BIE: the rate of occurrence, for example, the number of events per unit time. There is a large body of literature indicating that the organisms' relative distribution of behavior and time is controlled by the rates of occurrence of BIE (Davison & McCarthy, 1988; Herrnstein, 1961).

Additionally, we can consider a third statistical property of the distributions of BIE: the uncertainty about the time, place, covariances and rate of occurrence of BIE. If the uncertainty is expected, there is a probability distribution with parameters that can be inferred from experience, as in the case of "bandit" problem; if the uncertainty is unexpected, the parameters of the distribution change over time or space according to a transition function, for example, in bandit experiments where the reward probabilities of the arms change over time. This is a recent area of research with broad theoretical implications that we will see later in the chapter (Yu & Dayan, 2005).

A first conclusion we can reach is that if we accept the theory of natural selection and the statistical properties of the availability of BIE that can be described by the constraints stated above, we can say that psychology's natural object of study is the adaptability of behavior to the statistical properties of the environment.

In the recent years, the study of adaptive behavior has been advanced by the clarification of different levels of analysis (Marr, 1982). The first level is the computational or rational. It consists of studying of optimal solutions to problems posed by a particular set of constraints of the BIE. This level is closely related to an evolutionary analysis and to the way an engineer would approach the solution to a design problem. The analysis requires a detailed specification of the constraints involved. The difficulty in finding all the relevant restrictions to very specific behaviors makes this analysis more viable for problems with general constraints. A second level of analysis is called algorithmic and involves the study of the rules (mechanisms) that can implement the optimal solution. The third level is the biological implementation and searches for the physical structures that can implement a particular algorithm. The relation among levels is not one-to-one, but it is complementary. For example, different species can face similar problems of adaptation, and the computational analysis would converge to one solution. Different algorithms, however, can implement the solution depending on the elements available to compose a mechanism, and these can be implemented by different physical structures.

The way we described the relation between levels assumes that the computational one limits the other two, and it may seem more compatible with an evolutionary analysis. Nevertheless, the relation can go in the other direction: the physical subtract and the elements of the mechanisms can be seen as a limitation to what the "optimal" solution to a problem is. In conclusion, it is necessary to consider these three levels not only as mutual constraints, but also as complements. Typically, the solution to a problem is limited by the mechanisms and physical components available. Natural selection, however, can operate in a different direction promoting the evolution of new physical materials and mechanisms that underlie new solutions, as was the case with the evolution of the nervous system (Sterling & Laughlin, 2015).

Hill climbing and reinforcement learning are two good examples of integrating the three different levels of explanation. Consider first the problem of adaptation that salmonella faces in finding nutrients; it cannot perceive the nutrients at a distance. Nevertheless, experimental evidence indicates that the salmonella cluster around the attractant stimuli. To solve this problem, they have two behaviors: tumbling and straight swimming. A sudden change in the concentration of the nutrient produces a change in behavior from tumbling to straight swimming, which reverses to tumbling after a short period of time, and it changes again when there is a new change in the concentration of the nutrients. This behavior can be modeled with an algorithm that can generate an optimal solution to the problem of ending closer to the highest concentration of nutrients known as hill climbing.

Hill climbing is built on the following elements:

1. The possibility of detecting the element that is biologically relevant, in this case, a change in the concentration of the nutrient.

2. A memory of the value of that variable just a moment before.

3. A comparison of the value in that short-term memory with the current detected value.

4. Two behaviors: one behavior that randomly samples (random exploration) the value of the BIE, which is tumbling in the case of salmonella, and a second behavior that exploits (moves) in the direction of the improvement, which is straight swimming in the case of salmonella.

5. A rule that determines the value of the change in the relevant variable needed to change from a behavior of exploration to one of exploitation.

6. Very importantly, a process of adaptation to a change in the value of the relevant variable that allows returning to sampling after some time and escape from a local minima.

The integration of the six previous components can be succinctly captured by the following equation (Staddon, 2016):

$$Y_{t+1} = aY_t + b(X_{t+1} - X_t) \quad 0 < a < 1,$$

where Y is the concentration of the nutrient and represents a form of short-term memory; aY_t represents the factor of adaptation that refers to changes in the short-term memory; $b(X_{t+1} - X_t)$ represents the process of comparison between the level of concentration immediately before and the current level. The system has two responses and a threshold Y_0. If $Y \leq Y_0$, it continues tumbling (explore). If $Y > Y_0$, it changes to swimming straight (exploit). Note that in this simple model, there is no integration of the experiences: the bacteria only compare the experiences in two close points of time.

Reinforcement learning is another solution to the adaptive problem of detecting covariances among BIE and other events, including responses. These types of models are related to hill climbing (trial and error), however, they incorporate the possibility of integrating previous experiences. Reinforcement models are composed of two elements. The first one solves the problem of credit assignment (prediction), and the second is a response rule that determines the optimal use of the knowledge acquired. In what follows, we will only address the credit assignment problem by illustrating how the same mathematical structure can represent different intuitions. Computationally, we will assume that the goal of the system is to make the most accurate predictions; this will make it possible to maximize the reward rate. Later in the chapter we will discuss that depending on the statistical structure of the environment there may be other goals for the system.

Credit assignment is a computationally complex problem. It consists of determining the events responsible for the occurrence of a BIE, which in turn reduces the level of uncertainty associated with its presence. The main problem is the enormous space of possible candidates that can be credited for its occurrence. As a computational solution to the problem of credit assignment, reinforcement models require two

different steps. The first one is the reduction of the number of possible candidates, which can be accomplished through biases such as considering events that are contiguous, similar, novel or evolutionary relevant to the BIE. The second component is a mechanism that reduces through experience the number of candidates until only one element is left.

Perhaps the simpler representation of reinforcement learning is a leaky integrator. This model captures the early idea that reinforcement is a strengthening process where a reward "charges" a system that discharges in time if an additional reward is not provided. Bush and Mosteller (1953) formalized this class of models, which have dominated the theoretical and experimental literature in the study of learning until now (Dayan & Nakahara, 2018; Sutton & Barto, 1998).

We are interested in the dynamics, in discrete time, of the predictive value V_X of a stimulus X. The organism faces stimuli that unfold in time, sometimes alone and other times accompanied by a BIE. The leaky integrator states that the predictive value of a stimulus is a weighted sum of two variables, the predictive value of the stimulus at time t and whether a reward (BIE) might have occurred at time t or not:

$$Vx_{t+1} = aVx_t + (1-a)R_t \quad 0 < a < 1, \tag{1}$$

where the parameter a represents the importance of the cumulative experience with a stimulus X up to time t, relative to the occurrence or not of the BIE (R_t). Values of a close to one indicate that the accumulated experience is more important, while values close to zero indicate that the current presentation of a BIE is more relevant. The parameter a can be interpreted as a temporal window determining how far back in time an organism reaches for predicting the presentation of an BIE. In fact, a leaky integrator is an instance of an exponential running mean that gives more weight to the experiences closer to the present.

A simple rearrangement of the terms in equation (1) leads to a version of reinforcement learning more commonly used today, which is one that emphasizes a process of comparison:

$$Vx_{t+1} = Vx_t + \alpha(R_t - Vx_t). \tag{2}$$

If we set $\delta_t = \alpha(R_t - Vx_t)$, we have the equation better know as the δ rule:

$$V_{t+1} = V_t + \alpha\delta_t. \tag{3}$$

This form of the equation has had two different interpretations. One emphasizes the degree of “surprise” regarding the presence of an BIE. The second, more common today, emphasizes that the predictive value of a stimulus changes as a function of the magnitude of the prediction error and that the motor of learning is its reduction (Niv & Schoenbaum, 2008). In both forms of the equation, the computational goal is to achieve the best possible prediction; in one case as a filter that represents a running exponential mean, and in the other as an error reduction mechanism. The

findings that dopaminergic neurons fire in the presence of a prediction error is one of the best integration of the algorithmic and implementation levels (Shultz, Dayan, & Montague, 1997) and greatly support the study of reinforcement learning (Niv, 2009).

Equation (3), however, cannot account for the results of experimental protocols where two or more stimuli are present simultaneously and contiguous with an BIE, indicating that the stimuli compete for predictive value. These protocols are a more realistic representation of the problem of credit assignment. Rescorla and Wagner (1972) extended the standard reinforcement learning model to account for the results obtained, which suggests the following: one, compound stimuli are formed by separable elements with predictive value that are summed up, and two, that for each separated element of a compound stimuli the system applies the δ rule. If the compound stimuli consists of only two elements A and B, then

$$V_T = V_A + V_B, \tag{4}$$

where V_T represents the cumulative prediction. Thus, applying the δ rule for each element

$$V_{An+1} = V_{An} + \alpha(R_n - V_{T_n}). \tag{5}$$

The Rescorla-Wagner model has been successful in accounting for much empirical literature (Miller, Barnet, & Grahame, 1995) in complex credit assignment situations, and its simple mathematical structure has helped its acceptance in spite of known problems in accounting for phenomena such as spontaneous recovery of predictive value after periods of extinction (Bouton, 2018). Reinforcement learning algorithms are now an important tool in machine learning and robotics (Kober, Bagnell, & Peters, 2013). In addition, they have been extended to account for behavior in strategic interaction in the area of behavioral game theory (Camerer, 2003), in behavioral ecology (Frankenhuis, Panchanathan, & Barto, 2018), decision making (Erev & Haruvy, 2013) and in neuroeconomics (Daw, 2013).

In recent years, however, there have been new developments in reinforcement learning that allow representing adaptation in complex and changing environments that animals usually encounter in nature. For example, by incorporating the notion of uncertainty (Gershman, 2015; Kalman, 1960), long-term consequences (Sutton & Barto, 1998) and structural changes in the environment (Glaze, Filipowicz, Kable, Balasubramanian, & Gold, 2018; Nassar et al., 2012; Nassar, Wilson, Heasly, & Gold, 2010; Wilson, Nassar, & Gold, 2013). Under these circumstances, simple reinforcement learning algorithms like the one in equation (3) are unable to accurately describe behavior. This is particularly clear when studying behavior in unstable environments. Adaptation to these types of scenarios usually requires a concrete representation of change (Glaze et al., 2018; Ritz, Nassar, Frank, & Shenhav, 2018; Velázquez, Villarreal, & Bouzas, 2019) or parameters sensitive to the statistics of the environment (Nassar et al., 2010), neither of which is incorporated in equation (3).

Consider the case in which the reward rate in a foraging scenario gradually decreases because of continuous intake or increases because of the season of the year. These changes can occur at different speeds, which the animal can exploit to decide when is the best time to search for a different source. For example, if the reward rate is dangerously decreasing at a fast speed, it might be a better idea to start seeking other alternatives, even if the current level of reward is not too low. The idea that the animal can use information about the speed of change in the environment can be easily incorporated into equation (3):

$$V_{n+1} = V_n + V'_{n+1} + \alpha(R_n - V_n), \tag{6}$$

where V' represents both the direction (a rate of growth if positive and a rate of decrement if negative) and magnitude (the absolute value of V') of the change in the environment. This model is able to accurately describe behavior in environments that change at different rates (Velázquez et al., 2019) in perceptual decisions, and similar algorithms have been developed to describe reward-based decision (Kolling & Akam, 2017; Wittmann et al., 2016).

Importantly, changes in the environment can also be abrupt and unpredictable. In the last decade, an important body of research was developed tackling how humans (Nassar et al., 2012, 2010; Wilson et al., 2013) and other animals (Baum, 2010) adapt to these types of fluctuations. In the face of a change, past information becomes less relevant for making predictions about future outcomes. In the absence of changes, however, historical information becomes more relevant, since the future will likely resemble the past. The modulation of historical information is controlled by the learning rate α in equations (3) and (6), which should be dynamic in an environment with periods of stability and change. A simple extension to the structure of equation (6) allows the learning rate α to change depending on whether the environment is stable or has changed:

$$V_{n+1} = V_n + V'_{n+1} + \alpha_t(R_n - V_n) \qquad \alpha_t = \begin{cases} \alpha^{\text{fast}} & \text{if change} \\ \alpha^{\text{slow}} & \text{if no change} \end{cases}, \tag{7}$$

where the learning rate α now changes over time t (or trials) and can take one of two possible values. If the environment has changed, it becomes high, therefore allowing $fast$ adaptation. If the environment has not changed, it is low, which gives a higher relevance to historical information and promotes *slow* learning. Now, equation 8 represents a simple reinforcement learning algorithm that allows the adaptation to an environment that changes at a certain rate, which is represented by V', or that changes abruptly (by changing between fast or slow learning). Deciding between these two strategies of adaptation can easily be solved using Bayesian inference, which allows the selection of one or the other strategy depending on the current evidence in favor of either abrupt or gradual changes (following a rate) in the environment.

One of the main challenges in the application of mathematical models in psychology has been how to relate data and theory, given that as models get more complex, so do the methods required for data analysis. In this regard, the modeling of individual differences and model comparison have been the problems that have received the most attention. In recent years, the development of computational and numerical methods has made Bayesian methods more readily available. From these, Bayesian graphical models have seen a wide application in the Cognitive and Behavioral sciences. Fields like psychophysics (Lee, 2018), decision making (Scholten, Read, & Sanborn, 2014) and reinforcement learning (Velázquez et al., 2019), to name a few, have started to adopt the Bayesian methodology for making inferences about model parameters (Chávez, Villalobos, Baroja, & Bouzas, 2017), comparing models (Steingroever, Wetzels, & Wagenmakers, 2016) and designing experiments (Myung & Pitt, 2009; Zhang & Lee, 2010).

There are advantages that have come with the adoption of Bayesian graphical models in psychology. The first one is that these methods allow us to represent uncertainty in a principled way by means of a probability distribution over model parameters known as prior distribution. Second, this type of analysis makes it simpler to test for individual differences in a data set. Two recent examples can be found in the literature (Chávez et al., 2017; Villarreal et al., 2019) where these methods are used to show that individual differences that arise from patterns in the data have a common sense interpretation within relatively simple models. Third, the adoption of these methods has led to a wider use of Bayes Factors as a means of model comparison. This adoption has allowed researchers to avoid the problem of how to account for model complexity and overfitting, given that this method already deals with both problems. A good example of the application of Bayesian graphical models in psychology can be found in Lee (2018), where the author introduces the advantages mentioned here through a practical example with a psychophysics experiment. Another example of the advantages of a Bayesian approach to model construction and evaluation related to reinforcement learning can be found in a recent paper by Villarreal et al. (2019).

We can see from these simple examples that the nature of psychology and its relation to other sciences has changed substantially in the first quarter of this century. Experimental psychology has become a model-based science where computational and probabilistic models dominate, many of which are common in different scientific areas such as neuroscience, artificial intelligence, microeconomics and game theory. Additionally, the role played by statistics in psychology has changed. In particular, Bayesian statistics is now widely used for parameter estimation and model comparison. Along with these developments, the use of open software such as Python and R has facilitated replicability and open science. These developments urge us to rethink the structure of the academic programs aimed at the formation of behavioral scientists. They call for a new disciplinary degree in behavioral sciences different from the degree in professional psychology.

José Antonio de la Peña has been one of the greater driving forces of mathematics in México, constantly finding spaces where the language of mathematics can illuminate and help to clarify questions and solutions. In other chapters, you will find details of the richness of his life. However, we thought that the best tribute we can give him was to provide a couple of examples of how mathematics has helped the development of theories of learning in psychology. The conference tribute is a good example of another role that José Antonio has played in the advancement of the Science in México, well beyond his role in the leadership of scientific institutions and agencies that support them. His qualities as a human being make him a magnet that attracted a group of friends from very diverse areas of science, including psychology; dinners at his house turned into lively conversations from different perspectives, always with the amiability of Nelia Tello. We can all congratulate ourselves to have the friendship of José Antonio and his continuing support to integrate the work of different disciplines under the mantle of mathematics, as well as his teaching about how to face the inescapable uncertainty that surrounds our lives.

References

Batchelder, W., Colonius, H., Dzhafarov, E., & Myung, J. (Eds.). (2016). *New handbook of mathematical psychology* (Vol. 1-2). Cambridge: Cambridge University Press.

Baum, W. M. (2010). Dynamics of choice: A tutorial. *Journal of the Experimental Analysis of Behavior*, *94*(2), 161–174.

Bouton, M. E. (2016). *Learning and behavior : a contemporary synthesis.* Sunderland, MA, US: Sinauer Associates.

Bouton, M. E. (2018). Extinction of instrumental (operant) learning: interference, varieties of context, and mechanisms of contextual control. *Psychopharmacology*, 1-13. doi: https://doi.org/10.1007/s00213-018-5076-4

Bush, R. R., & Mosteller, F. (1953). A stochastic model with applications to learning. *The Annals of Mathematichal Statistics*, *24*, 559-585.

Camerer, C. F. (2003). *Behavioral game theory: Experiments in strategic interaction*. New York: Russell Sage Foundation.

Chávez, M. E., Villalobos, E., Baroja, J. L., & Bouzas, A. (2017). Hierarchical bayesian modeling of intertemporal choice. *Judgment and Decision Making*, *12*, 19-28.

Davison, M., & McCarthy, D. (1988). *The matching law: A research review*. Taylor and Francis Group.

Daw, N. (2013). Advanced reinforcement learning. In P. W. Glimcher & E. Fehr (Eds.), *Neuroeconomics: Decision making and the brain (2nd ed.)* (chap. 16). London, UK: Academic Pres.

Dayan, P., & Nakahara, H. (2018). Models and methods for reinforcement learning. In *Stevens' handbook of experimental psychology and cognitive neuroscience* (p. 1-40). American Cancer Society.

Erev, I., & Haruvy, E. (2013). Learning and the economics of small decisions. In J. H. Kagel & A. E. Roth (Eds.), *The handbook of experimental economics* (Vol. 2). Princeton, New Jersey: Princeton University Press.

Falmagne, J.-C., & Lee, M. D. (2002). Mathematical psychology. *International encyclopedia of the social and behavioral sciences*, 9405–9412.

Frankenhuis, W. E., Panchanathan, K., & Barto, A. G. (2018). Enriching behavioral ecology with reinforcement learning methods. *Behavioural Processes*.

Gershman, S. J. (2015). A unifying probabilistic view of associative learning. *PLoS computational biology*, *11*(11), e1004567.

Glaze, C. M., Filipowicz, A. L., Kable, J. W., Balasubramanian, V., & Gold, J. I. (2018). A bias–variance trade-off governs individual differences in on-line learning in an unpredictable environment. *Nature Human Behaviour*, *2*(3), 213.

Herrnstein, R. J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior*, *4*, 267–272.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, *82*, 35–45.

Kober, J., Bagnell, J. A., & Peters, J. (2013). Reinforcement learning in robotics: A survey. *The International Journal of Robotics Research*, *32*(11), 1238-1274.

Kolling, N., & Akam, T. (2017). (reinforcement?) learning to forage optimally. *Current opinion in neurobiology*, *46*, 162–169.

Lee, M. D. (2018). Bayesian methods in cognitive modeling. In E. J. Wagenmakers & J. T. Wixted (Eds.), *Stevens' handbook of experimental psychology and cognitive neuroscience (fourth edition)* (Vol. 5). John Wiley and Sons.

Luce, D. (1995). Four tensions concerning mathematical modeling in psychology. *Annual Review of Psychology*, *46*, 1-26.

Luce, R. D., Bush, R. R., & Galanter, E. (Eds.). (1963). *Handbook of mathematical psychology* (Vol. 1-3). Oxford, England: John Wiley.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York, NY, USA: Henry Holt and Co., Inc.

Miller, R. R., Barnet, R. C., & Grahame, N. J. (1995). Assessment of the rescorla-wagner model. *Psychological Bulletin*, *117*(3), 363-386.

Myung, J., & Pitt, M. (2009). Optimal experimental design for model discrimination. *Psychological Review*, *116*(3), 499-518.

Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., & Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature neuroscience*, *15*(7), 1040.

Nassar, M. R., Wilson, R. C., Heasly, B., & Gold, J. I. (2010). An approximately bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *Journal of Neuroscience*, *30*(37), 12366–12378.

Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology*, *53*(3), 139-154.

Niv, Y., & Schoenbaum, G. (2008). Dialogues on prediction errors. *Trends in Cognitive Science*, *12*(7), 265-272.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & P. W. F. (Eds.), *Classical conditioning II: Current research and theory* (p. 64-99). New York: Appleton-Century-Crofts.

Ritz, H., Nassar, M. R., Frank, M. J., & Shenhav, A. (2018). A control theoretic model of adaptive learning in dynamic environments. *Journal of Cognitive Neuroscience,*, *30*, 1405-1421.

Scholten, M., Read, D., & Sanborn, A. (2014). Weighing outcomes by time or against time? evaluation rules in intertemporal choice. *Cognitive Science*, *38*(3), 399-438.

Shultz, W., Dayan, P., & Montague, R. (1997). A neural substrate of prediction and reward. *Science*, *275*, 1593-1599.

Staddon, J. E. R. (2016). *Adaptive behavior and learning* (2nd ed.). Cambridge University Press.

Steingroever, H., Wetzels, R., & Wagenmakers, E.-J. (2016). Bayes factors for reinforcement-learning models of the iowa gambling task. *Decision*, *3*(2), 115-131.

Sterling, P., & Laughlin, S. (2015). *Principles of neural design*. The MIT Press.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. MIT press.

Velázquez, C., Villarreal, M., & Bouzas, A. (2019). Velocity estimation in reinforcement learning. *Computational Brain & Behavior*, 1–14.

Villarreal, M., Velazquez, C., Baroja, J. L., Segura, A., Bouzas, A., & Lee, M. (2019). Bayesian methods applied to the generalized matching law. *Journal of the Experimental Analysis of Behavior*.

Wilson, R. C., Nassar, M. R., & Gold, J. I. (2013). A mixture of delta-rules approximation to bayesian inference in change-point problems. *PLoS computational biology*, *9*(7), e1003150.

Wittmann, M. K., Kolling, N., Akaishi, R., Chau, B. K., Brown, J. W., Nelissen, N., & Rushworth, M. F. (2016). Predictive decision making driven by multiple time-linked reward representations in the anterior cingulate cortex. *Nature communications*, *7*, 12327.

Yu, A., & Dayan, P. (2005). Uncertainty, neuromodulation and attention. *Neuron*, *46*, 681-692.

Zhang, S., & Lee, M. (2010). Optimal experimental design for a class of bandit problems. *Journal of Mathematical Psychology*, *54*, 499-508.