What Have Computational Models Ever Done for Us?
A Case Study in Classical Conditioning

Eduardo Alonso, School of Informatics, City University London, London, UK & Centre for Computational and Animal Learning Research, St. Albans, UK

Esther Mondragón, Centre for Computational and Animal Learning Research, St. Albans, UK

ABSTRACT

The last 50 years have seen the progressive refinement of our understanding of the mechanisms of classical conditioning and this has resulted in the development of several influential theories that are able to explain with considerable precision a wide variety of experimental findings, and to make non-intuitive predictions that have been confirmed. This success has spurred the development of increasingly sophisticated models that encompass more complex phenomena. In such context, it is widely acknowledged that computational modeling plays a fundamental part. In this paper the authors analyze critically the role that computational models, as simulators and as psychological models by proxy, have played in this enterprise.

Keywords: *Classical Conditioning, Computational Models, Psychological Models, Psychology, Simulators*

INTRODUCTION

In natural environments organisms are compelled to constantly accommodate their behavior to dynamic surroundings. Learning to predict event regularities in such sensory rich conditions is vital for adaptive behavior and decision-making. Associative learning studies have mostly been conducted within the groundwork of *classical conditioning* –which is based on the principle that repeated pairings of two events will allow an individual to predict the occurrence of one of them upon presentation of the other, as consequence of the formation of an association between them (see Mackintosh, 1994; Wasserman & Miller, 1997; Pearce & Bouton, 2001). This simple mechanism is considered to underlie many learning phenomena and has proved to be relevant to human learning both theoretically (judgment of causality and categorization, e.g., (Shanks, 1995)) and practically, as the core of a large number of clinical models (Haselgrove & Hogarth, 2011; Schachtman & Reilly, 2011).

Hence, it is widely accepted that *classical conditioning is at the basis of most learning phenomena and behavior* and thus paramount that we develop accurate models of conditioning. In this endeavor, collaboration between psychologists and computer scientists has enjoyed considerable success (Schmajuk, 2010a; Schmajuk, 2010b; Alonso & Mondragón, 2011). This collaboration is sustained on well-known

DOI: 10.4018/ijalr.2014010101

arguments: Expressing models as sets of algorithms grants us formal ways of representing psychological intuitions and means of calculating their predictions accurately and quickly; from computational models we also borrow a view, the so-called computer metaphor, on how information is processed that has proved useful in understanding cognition; moreover, the architectures in which computational models are implemented, artificial neural networks for instance, resemble those of associative learning, both at conceptual and neural levels; finally, machine learning models, such as temporal difference learning and Bayesian learning, can be viewed as effective abstractions of how associations are formed and processed.

In this paper we analyze critically the assumptions upon which such arguments are built. We identify two main trends in so-called computational psychology, more in particular in the use of computational models in the study of conditioning, namely, as psychological models by proxy and as simulators.

COMPUTATIONAL MODELS AS MODELS OF LEARNING

Computational models of learning have been considered as psychological models in themselves. This position, that constitutes a milestone in the annals of cognitive science and artificial intelligence, is in fact a *misuse* of the term. We are illustrating our contention by means of a paradigmatic example, the use of Artificial Neural Networks (ANNs) in conditioning theory. In what follows we discuss the inadequacy of such approach at different levels of analysis, namely, ontological, formal, representational, functional, and structural.

The Ontological Level

ANNs are considered the substratum of conditioning. The motivational rationale is that (a) ANNs model by analogy natural neural networks and that (b) psychological processes, including conditioning, are ultimately embedded in natural neural networks; consequently, ANNs stand as a model of conditioning.

Despite the appeal in this line of argumentation, it is widely acknowledged that ANNs do not resemble natural neural networks in any fundamental way (Enquist & Ghirlanda, 2005); moreover, there is no strong evidence suggesting that neural activity and associative learning are indeed related (Morris, 1994) –or for that matter, that psychological processes can be uniquely localized in specific brain regions as recently shown in Vul, Harris, Winkielman, and Pashler (2009), and advanced in Uttal (2001).

Even if it did, a neural analysis would not necessarily shed light to the study of learning phenomena. In the words of B. F. Skinner "The analysis of behavior need not wait until brain science has done its part. The behavioral facts will not be changed (…). Brain science may discover other kinds of variables affecting behavior, but it will turn to a behavioral analysis for the clearest account of their effects" (Skinner, 1989, pp. 18). It should be noted that such a radical statement does not contradict a version of reductionism that most neuroscientists and cognitive psychologists would endorse, namely, Richard Dawkin's hierarchical reductionism (Dawkins, 1986), according to which one should determine the proper low explanatory level for the system under study.

The Formal Level

Some formalizations of ANNs and conditioning incorporate analogous formal descriptions. Nonetheless, that a version of Dirac's rule can be taken as a formal description of both neural plasticity and long-term potentiation effects –the Hebbian rule (Hebb, 1949)– and of association formation –as characterized by the Rescorla-Wagner rule (Rescorla & Wagner, 1972)– cannot be considered as proof of any common underlying structure and should not be used as an argument to reduce psychological phenomena to their alleged neural substratum.

Similarly, that the Rescorla-Wagner rule is essentially identical to the Widrow-Hoff rule (Widrow & Hoff, 1960) for training *Adeline* units and that, in turn, such rule can be seen as a primitive form of the generalized delta rule for backpropagation only tells us that,

computationally speaking, associative learning follows an error-correction algorithm[1]. A computational model does not identify however the underlying psychological factors (attention, motivation, etc.) involved in classical conditioning or how the physical characteristics of the elements (*e.g.*, the salience of the stimuli) affect such processes.

Clearly, sharing a common formal expression does not necessarily imply that the phenomena so expressed are of the same nature: Power functions, for example, can be used to express the relationship between (1) the orbital period of a planet and its orbital semi-major axis (Kepler's third law), (2) the metabolic rate of a species and their body mass (Kleiber's law), and (3) the magnitude of a stimulus and its perceived intensity (Stevens' law). To quote Richard Shull "The fact that an equation of a particular form describes a set of data does not mean that the assumptions that gave rise to the equation are supported. The same equation can be derived from very different sets of assumptions" (Shull, 1991, pp. 246).

In other words, we cannot assume that the meaning of a formal model is in the linguistic expression it takes (or that there is a unique isomorphism between phenomenon and algorithm). If we did, we would not be able to explain how a theory can be expressed in different sets of equations. Likewise, we would have no guarantees of the effect that the addition or the removal of a simple parameter may have. Paraphrasing (Chakravartty, 2001), theories and models can be given linguistic formulations but theories and models should not be identified with such formulations[2].

The Representational Level

ANNs are connectionist models that do not store information explicitly in symbols and rules but rather in the weights (strengths) of the connections. Following this interpretation, learning consists of changes in these weights. It is claimed, rightly, that this assumption underlies associative learning models and hence, wrongly, that ANNs are the substrate of associative learning phenomena. This quite straightforward argument is, in fact, a fallacy: As connectionists (at least implementational connectionists) themselves concede the way we represent learning in the network, either as continuous changes of weighted connections or as the result of discrete symbolic processing, is *a matter of convenience* and therefore irrelevant to the study of the structures involved.

This debate has been core in differentiating between "associative models" and "computational models" of learning: It is understood that associative models are historically and conceptually linked to connectionism (Medler, 1998) whereas computational approaches are inspired in the theory of information processing (Gallistel & Gibbon, 2001). Traditionally, the former approach is considered as sub-symbolic, therefore not formal, and the latter as symbolic, that is, computable by a Turing Machine. It can be argued however, following Peter R. Killeen, that both approaches are indeed formal (Killeen, 2001): Turing machines and ANNs are both computational models[3]; in particular, Turing Machines and ANNs are equivalent in their input/output behavior, that is, they compute the same problems and accept the same languages (Chomsky, 1956)[4].

The Functional Level

ANNs typically approximate solutions by iteratively minimizing an error function. A process that can be understood as a form of learning and that resembles learning by "trial and error", an instance of associative learning. However, it is worth emphasizing that ANNs merely implement numerical *methods*. They are, in fact, statistical tools –with a misleading name, and certainly not the simplest, fastest or most efficient techniques (see, *e.g.*, Mitchie, Spiegelhalter, & Taylor, 1994). On the other hand, associative learning models such as Rescorla and Wagner's express dynamic *laws*: Against public opinion, animals do not make predictions and iteratively update an associative value through error minimization towards an optimal one. The associative value at a given

time is the right associative value –that accurately describes the amount of learning at a given trial. Let's put it another way: In standard conditions, learning about the full extent of the association after one single pairing would be non-adaptive –except in situations in which there is preparedness bias for quick learning (Seligman, 1971). That the system described by Rescorla and Wagner's rule is limited by an asymptote (determined by the maximum reinforcing value of the US) does not confer any special status to such value –rather it just defines a constraint of the system.

The Structural Level

The layout of an ANN, the way units are connected within and between layers, can be interpreted as a cognitive architecture. Let's take a computational example to counter-argue this point: Network communications are designed and built following the Open Systems Interconnection model (OSI) (Zimmerman, 1980), from the physical layer that describes the electrical specifications of the devices the networks consist of up to the application layer that describes how the user interacts with a given piece of software. Thus, the OSI model implements a hierarchical and integrated architecture, that is, the type of cognitive architecture that a computational model of associative learning should allegedly support (Sun, 2008). Why don't we use the OSI model as a psychological model then? At the end of the day, structurally, OSI would make as good a psychological model as an ANN. That ANNs are networks implemented in architectures that take advantage of massive computational parallelism – not surprisingly, the *new connectionism* landmark paper introduced the Parallel Distributed Processing paradigm in cognition (Rumelhart & McClelland, 1986), does not confer them any psychological advantage: Any complex network would do (Newman, Barabási, & Watts, 2006).

TEMPORAL DIFFERENCE AS AN EXAMPLE

We are now narrowing down our analysis to a specific model of classical conditioning. Temporal Difference (TD) (Sutton & Barto, 1987; 1990) is an error-correction model of associative learning that has become popular due to the fact that it allegedly explains classical conditioning at the three Marr's levels: algorithmic, computational and physical. We are analyzing TD at each of these levels and arguing that regardless of which one is taken to model classical conditioning, TD does not succeed.

Algorithmic Level

From an algorithmic perspective, TD is a just a real-time extension of Rescorla and Wagner's model. As a consequence, TD inherits Rescorla and Wagner's limitations in predicting critical phenomena such as latent inhibition or spontaneous recovery. Current research trends in associative learning on the effects of attentional factors and on hierarchical structures are beyond the scope of this model. Even in dealing with temporal properties of conditioning, TD is severely restricted: First, the original TD model, in which stimuli are represented as single units regardless of their lengths, does not account for simple temporal discrimination. Its successor, CSC TD for Complete Serial Compounds TD, conceptualizes stimuli as a set of unique components (Moore, Choi, & Brunzell, 1998). In so doing, CSC TD faces the opposite problem that its predecessor, making it unable to predict temporal generalization. In addition, CSC TD is based on psychologically unrealistic assumptions such as the existence of a perfect clock. Recent attempts to solve these issues (microstimuli, (Ludvig, Sutton, & Kehoe, 2012)) are not exempt of problems. In summary, algorithmically TD is ill-equipped to predict phenomena it was explicitly designed for.

Computational Level

Computationally speaking, TD is not an accurate model of associative learning. Firstly, TD is based on the idea of optimization of a reward signal –animals are assumed to maximize reinforcement. Whereas this can be useful in control theory[5], it is by no means a universally accepted principle in studies of animal behavior and, in fact, contradicts empirical evidence (Staddon, 2007). Secondly, even if it were, TD does not converge to optimality in the general case (Fairbank & Alonso, 2012), and diverges in most biologically relevant problems (Ludvig, Bellemare, & Pearson, 2011).

Physical Level

TD's main appeal comes from describing associative learning at both behavioral and neural levels. Certainly, there seems to be a close correspondence between the behaviour of dopamine neurons in classical conditioning tasks and the prediction error in the TD algorithm (e.g., Schultz, Dayan, & Montague, 1997). How these studies advance our understanding of psychological processes is a different matter. At the end of the day, associative learning deals with psychological processes and psychological processes need to be explained at the psychological level. Other levels are not necessary or sufficient. The problem is that, in following Marr's analysis, TD considers the neural as the physical level and implicitly neglects the psychological level as an abstraction that is only useful in so much as it ultimately relates to "what really happens". This same reductionist régime could also be applied to the neural level displacing "what really happens" towards a molecular level, and thus regarding neuron behavior as a byproduct of the latter. Our position is that the physical level is the behavior of organisms, not neuronal spiking or blood flow.

Indeed, the confusion generated by using Marr's hypothesis in the analysis of classical conditioning has spurred extreme positions that consider purely psychological models as superfluous –in the words of C.R. Gallistel and Louis D. Matzel "(…) antirepresentational form of associative theorizing may need to be abandoned." (Gallistel & Matzel, 2013, pp. 174).

It is also argued that TD is a more comprehensive model of associative learning (than Rescorla and Wagner's for instance) in that it explains both classical conditioning and instrumental conditioning (that TD practitioners renamed as Reinforcement Learning, RL). TD advocates seem to forget that, psychologically speaking, the *associative structures* of classical and instrumental conditioning are the same (Hall, 2002). In both procedures, changes in behavior are considered the result of an association between two concurrent events and explained in terms of operations of a system that consists of nodes among which links can be formed. Moreover, a close inspection of the associations that RL explains reveals that it only covers one type of associations, arguably considered as an instance of instrumental learning: RL focuses its analysis on S-R associations, those originally described in early learning theory by Thorndike's law of effect to account for the formation of habits (Thorndike, 1898). RL does not operate in principle with other associations in instrumental conditioning such as R-O or S-R-O associations, without which goal-directed behavior cannot be addressed. This is an inherent flaw of RL, since the reinforcer is not considered as an element of the association but rather as a signal that in some way "stamps on the S-R association" (Alonso & Mondragón, 2013).

Alternative computational models based on Bayesian inference and so-called "rational" approaches have been proposed in the area with limited success. These theories introduce concepts and techniques alien to psychological theory – are not "penetrable" (Wills & Pothos, 2012) –, and fail dramatically when tested against classical conditioning phenomena (e.g., Gershman & Niv, 2012). Hence, it is difficult to evaluate how they can contribute to research in the field.

COMPUTATIONAL MODELS AS SIMULATORS OF MODEL OF LEARNING

A different way of analyzing computational modeling in classical conditioning predicates that computational models can be considered as implementations of psychological models – rather than as psychological models themselves. In this sense, a computational model is a tool to generate simulations that serve two main purposes: On the one hand, implementing a model requires precise definitions, in the manner of a formal model or specific programming language instructions, which make the pre-existing psychological model "accountable". On the other hand, algorithms empower us to execute calculations rapidly and, most importantly, accurately. Automation is paramount, particularly when the models involve non-linear equations that can only be solved numerically as it is the case of recent models of conditioning (Balkenius & Morén, 1998; Vogel, Castro, & Saavedra, 2004; Mitchell & Le Pelley, 2010; Schmajuk & Alonso, 2012). Crucially, the outputs of a simulation feedback the psychological models –thus becoming cardinal in the cycle of theory formation and refinement.

Nonetheless, the advantages derived from using implementations do not spring solely from the formal specification of the psychological models in equations and algorithms. Per se, such descriptions constitute a mathematical model, a necessary yet no sufficient condition for a formal model to be computational. The essence of a computational model lies in the fact that it is implemented. From this view, in psychology, as well as in other empirical sciences, a computational model is a model that has been simulated. In linguistic terms, we need to add semantic and pragmatic content to the syntactic description of the model for it to become computational.

This interpretation is not without critics: It has been argued that a model is computational if it is "implementable" –regardless of whether it is in fact described as a full-bodied computational model. This point of view can be considered as an *abuse* of the term "computational" since any psychological model of conditioning would fit this definition. Stretching the analogy, this use of the term "computational" would make all models of physics since Galileo's computational.

From a theoretical point of view a computational model might be interpreted as a mere formalization of the concept of computation, which in turn does not necessarily require its implementation in a computer. Mathematically, the notion of computation is a mechanical or automated procedure, an algorithm (Turing, 1937). Computers are physical instantiations of the abstract machines that would compute such procedures. In fact, this definition was proposed well before the first digital general-purpose computers had even been designed. Contrarily, our standpoint is that a computational model needs to be implemented in a computer if it is to add further to what constitutes a mathematical model in its own right.

An alternative source of disagreement about the use of the term computational can potentially be traced to cognitive science. As introduced in our discussion of TD above, the term "computational" has been linked to David Marr's Tri-Level Hypothesis on vision. In Marr's theory the "what" level translates as the computational level, the "how" refers to the algorithmic level and the "where" to the implementational level (Marr, 1982). However insightful such analysis may be, clearly what Marr denoted as "computational" refers to the psychological process itself –when applied to cognition. Insisting on talking about psychological models as if they were computational based on such taxonomy is, in our opinion, a source of misunderstanding.

MODEL SELECTION

Debating the characteristics of a good computational model of psychology influences the selection of models and in turn may help us determine what a computational model "truly" is.

The selection of a model, psychological or otherwise, described in natural language or

mathematically, is not an easy task. It relies in formal definitions and methods as well as on scientific practice and common sense (Kuhn, 1962; Feyerabend, 1975). Although quantitative formulas have been developed to compare models, based on the average size of the deviations from predicted values, the number of data points and free parameters (Akaike, 1974; Schwarz, 1978), relying exclusively on such formalisms or applying blindly Occam's razor is not advisable. According to Baum (1983) evaluating a model requires good judgment based on careful consideration of many factors, both technical and logical. The very essence of a model reflects the choices scientists make –what they consider relevant beyond the mere quantitative. It is thus critical that the community reaches an agreement on how to evaluate and compare their models. Unfortunately, it is often the case that researchers in a given area focus on small datasets that don't cross domains. This conveys a lack of consensus on critical phenomena, on whether the number of parameters should be "penalized" or on whether the parameters should be fixed or optimized for each condition (Alonso & Schmajuk, 2012).

Consequently, we are compelled to question what to assess when evaluating a computational model of learning.

If computational models are simulators then we would need to select amongst them according to their computational complexity, which is related to but not reducible to the algorithms they implement. In other words, time and space (memory) of computation becomes paramount in the decision. In addition, these computing tools must be tested for reliability and dependability against failures –which, in turn, depend on various factors such as programming languages, operating systems, memory capacity, processing speed, as well as on software engineering and management requirements. Computational models as simulators add a new level of sophistication. But this sophistication comes at a price: A computer program is not as "aseptic" as a mathematical description. A computer program takes life in algorithms and data structures that must comply with software and hardware specifications.

Regrettably, the state of the art in simulation of classical conditioning is not very encouraging: Although classical conditioning software has been recently described in the literature (Schultheis, Thorwart, & Lachnit, 2008a; Schultheis, Thorwart, & Lachnit, 2008b; Thorwart, Schulthei, König, & Lachnit, 2009; Alonso, Mondragón, & Fernández, 2012; Mondragón, Alonso, Fernández, & Gray, 2013; Mondragón, Gray, & Alonso, 2013), it is still the case that most psychologists in the area view simulations as mere addenda rather than as an integral part of experimental methodology. Simulation software is developed, implemented and documented in an ad hoc manner, raising serious concerns about its reliability, usability and scalability. As such, their impact is very limited –which conflicts with the widely accepted opinion that simulations are decisive in the development of accurate models of classical conditioning.

On the other hand, if computational models are considered as a valid alternative to psychological models, which criteria should be used to evaluate them and choose amongst them? There is no a clear answer to this question.

PHILOSOPHICAL ISSUES

A final more general reason to explain the appeal of computational models in psychology rests on the idea of isomorphism between computers and the brain, rebranding them as information processing systems, instantiations of a universal Turing machine or any other model of computation. But this idea alone does not justify the vast support that the "computer metaphor" enjoys. After all, any phenomena can potentially be expressed in terms of some sort of computation. The reason why this is such a powerful metaphor lies of the fact that this line of rationalization it is deeply rooted in Western philosophy and the mechanization of (formal) reasoning, reformulated in the twentieth century in terms of computation. That

computation has been effectively embedded in computers has reinforced the idea that the same must be true is in the brain, that the study of the former will help understand the latter and, in a *tour the force*, that computers may be capable of displaying intelligence. Indeed, every scientific theory is shaped in the context of its age's achievements and prejudices: Like Newton's laws of mechanics strengthened the view of the Universe as a deterministic machine that worked as the sophisticated clocks so popular at the time, our conception of the mind as an information processing machine à la Turing has certainly been influenced by the development of computing technology.

And precisely because of its generality the information processing model is of little use: Working physicists do not model electrons, atoms or galaxies as information processing entities –be it in the form of a cellular automaton as envisaged in Zuse (1969) or as a participatory universe (Wheeler, 1990); on the other hand, neither (computational) physicists nor the public would presume that the simulation of a nuclear reaction generates real energy or that a flight simulator really flies. Of course, this does not preclude physicists from theorizing about what type of information is contained in a physical system (see, for example, literature on quantum entanglement or black holes) or about exploring the physical limits of computers (pioneered by Richard Feynman (Hey & Allen, 2000), and followed up to contemporary theories of quantum computing e.g., Vedral, 2006). But these debates are not part of mainstream physics.

CONCLUSION

To sum it up, although the need to get influx from "outsiders" is recognized within the psychology community computational models should be taken with caution (see Townsend, 2008). Computational models may provide us with complementary idealized models of psychological phenomena; they can also offer powerful statistical tools upon which psychologists can build data models; but computational models alone are not the appropriate methods to answer psychological questions. This is an obvious, hardly original, conclusion –and yet more often than not we read flamboyant news about robots that learn, think and experience emotions or ANNs that can do anything psychological models do only better. On the other hand, given the increasing complexity of psychological models of conditioning developing accurate and rapid simulators to test their predictions is, in our opinion, a must.

We would like to conclude with two warnings against extreme cases in the misuse of computational models as psychological models. The first goes like this: We take psychological data and write a program that fits it. Since the data is psychological the program must constitute a psychological model. It should be obvious, however, that curve fitting does not automatically make a model "psychological" (or "biological" or "physical"). It must provide psychological insight. The second one is a cautionary note against hype: As an illustration, simple programs that, to some extent, learn to maximize a numerical signal by trial and error have been presented as a "theory of mind" (Sutton, 2003). The history of Artificial Intelligence should have taught us better.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. doi:10.1109/TAC.1974.1100705

Alonso, E., & Mondragón, E. (2011). *Computational neuroscience for advancing artificial intelligence: Models, methods and applications*. Hershey, PA: Medical Information Science Reference.

Alonso, E., & Mondragón, E. (2013, February 15-18). Associative reinforcement learning: A proposal to build truly adaptive agents and multi-agent systems. In J. Filipe & A. Fred (Eds.), *Proceedings of the Fifth International Conference on Agents and Artificial Intelligence (ICAART 2013)* (Vol. 1, pp. 141-146). Barcelona, Spain: SciTe Press.

Alonso, E., Mondragón, E., & Fernández, A. (2012). A Java simulator of Rescorla and Wagner's prediction error model and configural cue extensions. *Computer Methods and Programs in Biomedicine*, *108*, 346–355. doi:10.1016/j.cmpb.2012.02.004 PMID:22420931

Alonso, E., & Schmajuk, N. (2012). Computational models of classical conditioning guest editors' introduction. *Learning & Behavior*, *40*, 231–240. doi:10.3758/s13420-012-0081-7 PMID:22926998

Balkenius, C., & Morén, J. (1998). Computational models of classical conditioning: A comparative study. In J.-A. Mayer, H. L. Roitblat, S. W. Wilson, & B. Blumberg (Eds.), *From animals to Animats 5*. Cambridge, MA: MIT Press.

Baum, W. M. (1983). Matching, statistics, and common sense. *Journal of the Experimental Analysis of Behavior*, *39*, 499–501. doi:10.1901/jeab.1983.39-499 PMID:16812332

Chakravartty, A. (2001). The semantic or model-theoretic view of theories and scientific realism. *Synthese*, *127*, 325–345. doi:10.1023/A:1010359521312

Chomsky, N. (1956). Three models for the description of language. *I.R.E. Transactions on Information Theory*, *2*, 113–124. doi:10.1109/TIT.1956.1056813

Dawkins, R. (1986). *The blind watchmaker*. New York, NY: W. W. Norton & Company.

Dirac, P. A. M. (1938). The relation between mathematics and physics. In *Proceedings of the Royal Society,* Edinburgh, UK (Vol. 59, 1938-39, Part II, 122-129).

Enquist, M., & Ghirlanda, S. (2005). *Neural networks & animal behavior*. Princeton, NJ: Princeton University Press.

Fairbank, M., & Alonso, E. (2012). The divergence of reinforcement learning algorithms with value-iteration and function approximation. In *Proceedings of the IEEE International Joint Conference on Neural Networks 2012 (IJCNN'12)* (pp. 3070–3077). IEEE Press.

Feyerabend, P. (1975). *Against method*. New York, NY: Verso Books.

Gallistel, C. R., & Gibbon, J. (2001). Computational versus associative models of simple conditioning. *Current Directions in Psychological Science*, *10*(4), 146–150. doi:10.1111/1467-8721.00136

Gallistel, C. R., & Matzel, L. D. (2013). The neuroscience of learning: beyond the Hebbian synapse. *Annual Review of Psychology*, *64*, 169–200. doi:10.1146/annurev-psych-113011-143807 PMID:22804775

Gershman, S. J., & Niv, Y. (2012). Exploring a latent cause model of classical conditioning. *Learning & Behavior*, *40*, 255–268. doi:10.3758/s13420-012-0080-8 PMID:22927000

Hall, G. (2002). Associative structures in Pavlovian and instrumental conditioning. In H. Pashler, S. Yantis, D. Medin, R. Gallistel, & J. Wixted (Eds.), *Stevens' handbook of experimental psychology* (Vol. 3, pp. 1–45). Hoboken, NJ: John Wiley and Sons. doi:10.1002/0471214426.pas0301

Haselgrove, M., & Hogarth, L. (2011). *Clinical applications of learning theory*. London, UK: Psychology Press.

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. New York, NY: Wiley.

Hey, T., & Allen, R. W. (Eds.). (2000). *Feynman lectures on computation*. Boulder, CO: Westview Press.

Killeen, P. R. (2001). The four causes of behavior. *Current Directions in Psychological Science*, *10*(4), 136–140. doi:10.1111/1467-8721.00134 PMID:19081757

Kuhn, T. (1962). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.

Ludvig, E. A., Bellemare, M. G., & Pearson, K. G. (2011). A primer on reinforcement learning in the brain: Psychological, computational, and neural perspectives. In Alonso & Mondragón (Eds.), Computational neuroscience for advancing artificial intelligence: Models, methods and applications (pp. 111-144). Hershey, PA: IGI Global.

Ludvig, E. A., Sutton, R. S., & Kehoe, E. J. (2012). Evaluating the TD model of classical conditioning. In E. Alonso and N. Schmajuk (Eds.), Special Issue on Computational Models of Classical Conditioning, Learning & Behavior, 40, 305-319.

Mackintosh, N. J. (Ed.). (1994). *Animal learning and cognition*. San Diego, CA: Academic Press.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco, CA: W.H. Freeman.

McCulloch, W., & Pitts, W. (1943). A logical calculus of ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, *5*, 115–133. doi:10.1007/BF02478259

Medler, D. A. (1998). A brief history of connectionism. *Neural Computing Surveys*, *1*(2), 18–73.

Mitchell, C., & Le Pelley, M. (2010). *Attention and associative learning: From brain to behaviour*. Oxford, UK: OUP.

Mitchie, D., Spiegelhalter, D. J., & Taylor, C. C. (Eds.). Elder, J. F., IV. (Rev.). (1994). Machine learning, neural, and statistical classification. Journal of the American Statistical Association, 91, 436-438.

Mondragón, E., Alonso, E., Fernández, A., & Gray, J. (2013). A Rescorla and Wagner simulator with context compounds. *Computer Methods and Programs in Biomedicine*, *110*(2), 226–230. doi:10.1016/j.cmpb.2013.01.016 PMID:23453075

Mondragón, E., Gray, J., & Alonso, E. (2013). A complete serial compound temporal difference simulator for compound stimuli, configural cues and context representation. *Neuroinformatics*, *11*(2), 259–261. doi:10.1007/s12021-012-9172-z PMID:23161265

Moore, J., Choi, J., & Brunzell, D. (1998). Predictive timing under temporal uncertainty: The TD model of the conditioned response. In D. Rosenbaum, & A. Collyer (Eds.), *Timing of behavior: Neural, computational, and psychological perspectives* (pp. 3–34). Cambridge, MA: MIT Press.

Morris, R. G. M. (1994). The neural basis of learning with particular reference to the role of synaptic plasticity: Where are we a century after Cajal's speculations? In N. J. Mackintosh (Ed.), *Animal learning and cognition* (pp. 135–183). San Diego, CA: Academic Press. doi:10.1016/B978-0-08-057169-0.50012-7

Newman, M., Barabási, A.-L., & Watts, D. J. (2006). *The structure and dynamics of networks*. Princeton, NJ: Princeton University Press.

Orponen, P. (1994). Computational complexity of neural networks: A survey. *Nordic Journal of Computing*, *1*(1), 94–110.

Pearce, J. M., & Bouton, M. E. (2001). Theories of associative learning in animals. *Annual Review of Psychology*, *52*, 111–139. doi:10.1146/annurev.psych.52.1.111 PMID:11148301

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: The effectiveness of reinforcement and non-reinforcement. In A. H. Black, & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York, NY: Appleton-Century-Crofts.

Rumelhart, D. E., & McClelland, J. L.PDP Research Group (Eds.). (1986). Parallel distributed processing: Vol. 1. *Foundations*. Cambridge, MA: The MIT Press.

Schachtman, T. R., & Reilly, S. (2011). *Associative learning and conditioning theory: Human and non-human applications*. Oxford, UK: Oxford University Press. doi:10.1093/acprof:oso/9780199735969.001.0001

Schmajuk, N. A. (2010a). *Computational models of conditioning*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511760402

Schmajuk, N. A. (2010b). *Mechanisms in classical conditioning: A computational approach*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511711831

Schmajuk, N. A., & Alonso, E. (2012). Special issue on computational models of conditioning. *Learning & Behavior*, *40*(3).

Schultheis, H., Thorwart, A., & Lachnit, H. (2008a). HMS: A MATLAB simulator of the Harris model of associative learning. *Behavior Research Methods*, *40*, 442–449. doi:10.3758/BRM.40.2.442 PMID:18522054

Schultheis, H., Thorwart, A., & Lachnit, H. (2008b). Rapid-REM: A MATLAB simulator of the replaced elements model. *Behavior Research Methods*, *40*, 435–441. doi:10.3758/BRM.40.2.435 PMID:18522053

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*, 1593–1599. doi:10.1126/science.275.5306.1593 PMID:9054347

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464. doi:10.1214/aos/1176344136

Seligman, M. E. P. (1971). Phobias and preparedness. *Behavior Therapy*, *2*, 307–321. doi:10.1016/S0005-7894(71)80064-3

Shanks, D. R. (1995). *The psychology of associative learning*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511623288

Shull, R. L. (1991). Mathematical description of operant behavior: An introduction. In I. H. Iversen, & K. A. Lattal (Eds.), *Experimental analysis of behavior* (Vol. 2, pp. 243–282). New York, NY: Elsevier. doi:10.1016/B978-0-444-81251-3.50014-X

Skinner, B. F. (1989). The origins of cognitive thought. *The American Psychologist*, *44*(1), 13–18. doi:10.1037/0003-066X.44.1.13

Staddon, J. E. (2007). Is animal behavior optimal? In A. Bejan, & G. W. Merkx (Eds.), *Constructal theory of social dynamics*. New York, NY: Springer. doi:10.1007/978-0-387-47681-0_8

Sun, R. (2008). *The Cambridge handbook of computational psychology*. Cambridge University Press. doi:10.1017/CBO9780511816772

Sutton, R. A. (2003). *Reinforcement learning's computational theory of mind*. Retrieved from http://webdocs.cs.ualberta.ca/~sutton/Talks/Talks.html

Sutton, R. S., & Barto, A. G. (1987). A temporal-difference model of classical conditioning. In *Proceedings of the Ninth Annual Conference of the Cognitive Science Society* (pp. 355-378).

Sutton, R. S., & Barto, A. G. (1990). Time-derivative models of Pavlovian reinforcement. In M. Gabriel, & J. W. Moore (Eds.), *Learning and computational neuroscience* (pp. 497–537). Cambridge, MA: MIT Press.

Thorndike, E. L. (1898). *Animal intelligence: An experimental study of the associative processes in animals*. PhD Dissertation, Columbia University, New York, NY.

Thorwart, A., Schultheis, H., König, S., & Lachnit, H. (2009). ALTSim: A MATLAB simulator for current associative learning theories. *Behavior Research Methods*, *41*(1), 29–34. doi:10.3758/BRM.41.1.29 PMID:19182121

Townsend, J. T. (2008). Mathematical psychology: Prospects for the 21st century: A guest editorial. *Journal of Mathematical Psychology*, *52*, 269–280. doi:10.1016/j.jmp.2008.05.001 PMID:19802342

Turing, A. M. (1937). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, *2*(42), 230–265. doi:10.1112/plms/s2-42.1.230

Turing, A. M. (1948). *Intelligent machinery*. Retrieved from http://www.turingarchive.org/browse.php/C/11.

Uttal, W. R. (2001). *The new phrenology: The limits of localizing cognitive processes in the brain*. Cambridge, MA: The MIT Press.

Vedral, V. (2006). *Introduction to quantum information science*. Oxford, UK: Oxford University Press. doi:10.1093/acprof:oso/9780199215706.001.0001

Vogel, E. H., Castro, M. E., & Saavedra, M. A. (2004). Quantitative models of Pavlovian conditioning. *Brain Research Bulletin*, *63*, 173–202. doi:10.1016/j.brainresbull.2004.01.005 PMID:15145138

Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly high correlations in fRMI studies of emotion, personality, and social sciences. *Perspectives on Psychological Science*, *4*(3), 274–290. doi:10.1111/j.1745-6924.2009.01125.x

Wasserman, E. A., & Miller, R. R. (1997). What's elementary about associative learning? *Annual Review of Psychology*, *48*, 573–607. doi:10.1146/annurev.psych.48.1.573 PMID:9046569

Werbos, P. J. (1974). *Beyond regression: New tools for prediction and analysis in the behavioral sciences*. PhD dissertation, Harvard University, Cambridge, MA.

Werbos, P. J. (1977). Advanced forecasting for global crisis warning and models of intelligence. In General systems yearbook.

Wheeler, J. A. (1990). Information, physics, quantum: The search for links. In W. Zurek (Ed.), *Complexity, entropy, and the physics of information*. Redwood City, CA: Addison-Wesley.

Widrow, G., & Hoff, M. E. (1960). *Adaptive switching circuits. Institute of radio engineers*. Western Electronic show & convention, Convention record, Part 4, 96-104.

Wills, A. J., & Pothos, E. M. (2012). On the adequacy of current empirical evaluations of formal models of categorization. *Psychological Bulletin*, *138*, 102–125. doi:10.1037/a0025715 PMID:22061692

Zimmerman, H. (1980). OSI reference model – The ISO model of architecture for open systems interconnection. *IEEE Transactions on Communications*, *28*(4), 425–432. doi:10.1109/TCOM.1980.1094702

Zuse, K. (1969). *Rechnender Raum*. Braunschweig: Friedrich Vieweg & Sohn. doi:10.1007/978-3-663-02723-2

ENDNOTES

1 Incidentally, backpropagation is merely a mathematical procedure to deriving partial derivatives –originally proposed to model social interactions not neural networks (Werbos, 1974).

2 The ontological properties of mathematic constructs have been historically considered by scientists and philosophers alike (Dirac, 1938). However, to the best of our knowledge that scientific propositions must be expressed mathematically does not confer them a special

relation with the phenomena under study. In other words, Descartes' ontological argument does not seem to be an appropriate scientific method.

[3] It should be noted that the first mathematical models of (A)NNs, McCulloch and Pitts's (McCulloch & Pitts, 1943) and Turing's B-type machines (Turing, 1948) were intended to formalize logically, i.e., symbolically, the notion of learning.

[4] Provided that the values of the weights are restricted to rational numbers (Orponen, 1994).

[5] TD was originally developed in control theory by Paul Werbos (Werbos, 1977), and coined as Heuristic Dynamic Programming.