Computational Models of Classical Conditioning
A Qualitative Evaluation and Comparison

Eduardo Alonso[1,2], Pavandeep Sahota[1] and Esther Mondragón[2]
[1]*Department of Computer Science, City University London, London EC1V 0HB, U.K.*
[2]*Centre for Computational and Animal Research Centre, St. Albans AL1 1RQ, U.K.*

Keywords: Computational Models, Classical Conditioning, Quantitative and Qualitative Evaluation, Comparison.

Abstract: Classical conditioning is a fundamental paradigm in the study of learning and thus in understanding cognitive processes and behaviour, for which we need comprehensive and accurate models. This paper aims at evaluating and comparing a collection of influential computational models of classical conditioning by analysing the models themselves and against one another qualitatively. The results will clarify the state of the art in the area and help develop a standard model of classical conditioning.

1 INTRODUCTION

In natural environments, there is a constant need for organisms to accommodate their behaviour to dynamic surroundings. Learning to predict the regularities in such sensory rich conditions is the key for adaptive behaviour and decision-making. Predictive learning studies have mostly been conducted within the context 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 a link between them (see Mackintosh, 1994; Pearce and Bouton, 2001; Hall, 2002). This simple idea is at the basis of many associative 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 good number of clinical models (Haselgrove and Hogarth, 2011; Schachtman and Reilly, 2011).

The last 50 years has 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 modelling plays a fundamental part (e.g., Dayan and Abbot, 2001; Schmajuk, 1997; 2010a).

There are two main motivations for using computational models: on the one hand, be it in the form of a specific programming language or as a formal model, implementations require unambiguous definitions that make the underlying psychological models more precise. On the other hand, algorithms allow us to execute calculations rapidly and, most importantly, accurately. The outputs of a simulation feedback the psychological models –thus becoming an essential part of the cycle of theory formation and refinement. Automation is critical, particularly when models are described in non-linear equations that can only be solved numerically as it is the case of recent models of conditioning (Vogel et al., 2004; Schmajuk, 2010b; Alonso and Mondragón, 2011). In particular, (Schmajuk and Alonso, 2012) brought together as a special issue on computational models of classical conditoining a collection of papers that represent the leading edge of the field. Henceforth we are referring to the papers in the issue by acronysms of the models themselves or the by the initials of the authors if none was given, that is, we are coining them GP, LCT, SLGK, PHK+, TD, MKM/APECS, AMAN and SOCR, respectively. Notwithstanding the relative merits of each model, as a theoretical corpus (Schmajuk and Alonso, 2012) showed that there is no unanimity on what the basic principles

Alonso E., Sahota P. and Mondragón E..
Computational Models of Classical Conditioning - A Qualitative Evaluation and Comparison.
DOI: 10.5220/0004903105440547
In *Proceedings of the 6th International Conference on Agents and Artificial Intelligence* (ICAART-2014), pages 544-547
ISBN: 978-989-758-015-4

and mechanisms of classical conditioning are or on standard procedures to investigate them. Although there is agreement, or at least some convergence, that learning is driven by the minimization of prediction error (but see Witnauer et al. above for a different view), the models considered differ substantially on the nature of stimulus representation (configural vs. elemental), the role of attention in the formation of associations, and about how temporal properties affect conditioning.

In order to build more comprehensive theories of classical conditioning it is thus critical that we carry out an exhaustive analysis of such models, that is, that we evaluate them and compare them against one another. Crucially, three requirements for contributors to the special issue were set (Alonso and Schmajuk, 2012): (1) models should be tested against a list of phenomena for which there was a consensus about their reliability; (2) model parameters should be fixed across simulations; and (3) authors should make available the simulations they used to test their models. In short, the models and their simulations should be replicable.

The list of phenomena was compiled by *domains*, as follows: acquisition phenomena (6 phenomena), extinction (3), generalization (3), discriminations (17), inhibitory conditioning (6), combination of separately trained CSs (3), stimulus competition/potentiation in training (11), CS/US preexposure effects (11), transfer (4), recovery (8), higher-order conditioning (5), and temporal properties (9). Phenomena were characterised as “General”, meaning that results had been demonstrated in a wide variety of procedures/organisms, or “Some Data” otherwise.

Regardless of the advances reported, (Schmajuk and Alonso, 2012) demontrated that models in the area are still partial (no model covers all the phenomena under investigation), incomplete (there are phenomena unaccounted for) and to some extent inconsistent (different models make contradictory predictions). (Schmajuk and Alonso, 2012) represents the vanguard in computational models of classical conditioning and, at the same time, provides us with the appropriate tools to evaluate and compare them.

2 EVALUATION

The over-reaching goal of this position paper is to diagnose the state of the art in computational modelling of classical conditioning, explain divergences and convergences, and identify those models that seem more promising in the search for a standard model of classical conditioning.

The evalution consists of two phases: a preliminary analysis of the software used in each case. Additionally, we are also considering how intuitive the underlying psychological assumptions of each model are, and other factors such as how many domains of phenomena each model crosses, that is, their generality, and whether they account for critical phenomena (for instance, latent inhibition or spontaneous recovery). Before proceeding, it should be noted that by a “computational model” we mean an implementation of a (pre-existing) psychological model, that is, we don’t consider computational models as formal models that act as psychological models by proxy. Also, we do not enter into the philosophical debate about the different levels at which psychological phenomena can be interpreted and about the relationship between the so-called computational level and other levels, algorithmic or physical (see, (Alonso and Mondragón, 2012) for a review on the uses, abuses and misuses of the concept “computational” in psychology).

2.1 Software

It is beyond the purpose of this paper to carry out validation and verification tests on the simulators in wich the computational models in (Schmajuk and Alonso, 2012) were run. We are not checking the replicability of the results reported either. Instead, we are summarizing, Table 1, which programming language was used in each case, whether it was documented (including a user’s guide), and whether the code was made available.

Table 1: Software.

Model	Language	Document	Code	Guide
SLGK	C	Y	Y	Y
AMAN	MATLAB	Y	Y	Y
GP	MATLAB	Y	Y	Y
PKH+	Visual Basic	Y	Y	Y
TD	MATLAB	N	Y	N
LCT	MATLAB	N	Y	N
MKM/AMEC	MATLAB	N	N	N
SOCR	MATLAB	N	N	N

It is up to the reader to decide whether, given the resources made available to them by the authors, the results reported are trustworthy. We are only commenting on the programming language used and on the software development characteristics that underlies all simulators. Regarding the former, MATLAB was the preferred choice. From the point of view of a programmer, MATLAB is relatively easy to learn and to use (at least, for simple applications). Speed-wise MATLAB is rather similar to alternatives like C, no matter whether they compile or interpret. One of MATLAB's disadvantages is that it is not a fully bodied programming language, and the user is not able to create modular programs and reusable code with it.

In addition, MATLAB is proprietary software and a proprietary language. MATLAB works only with MathWork's MATLAB software – meaning that if you have created programs in MATLAB, you will generally only be able to use those programs in MATLAB, and would need to do extensive porting to move to a different platform. MATLAB is not a platform-independent language.

More generally, most simulators are not professionally developed, failing to address the following issues:

- Inputting data is cumbersome.
- The system must be run afresh each time the input parameters are changed.
- Outputs cannot be directly exported and manipulated in widespread data processors such as, for example, excel.
- Interfaces and visualization of data are poor.
- Simulators are not portable across platforms.
- Simulators cannot be scaled up to accommodate new parameters and/or models.

Although classical conditioning software has been recently described in the literature (Schultheis et al., 2008a; 2008b; Thorwart et al., 2009; Alonso et al., 2012; Mondragón et al., 2013a; 2013b), it is still the case that most psychologists in the area view simulations as mere tools rather than as an integral part of experimental methodology. Software is developed, implemented and documented in an ad hoc manner, raising serious concerns about its reliability, usability and scalability.

2.2 Qualitative Analysis

The very essence of a model refers to the choices scientists make –choices that reflect what they consider relevant– and thus evaluating a model requires careful consideration of many factors, both technical and formal (Baum, 1983). However, in assessing and selecting models (and in identifying which features a good model should show) it is critical that we use measurable criteria (see (Shiffrin et al., 2008) for a recent survey). Typically, the behaviour of a model is considered locally, that is, at its best fitting parameter values. This approach is problematic, since best fits leave us with snapshots of the model's performance that are difficult to piece together into a comprehensive, global understanding of the model. In addition, quantitative analysis based on goodness-to-fit criteria can result in selecting overly complex models that generalize poorly. Finally, comparing models is even more difficult with local quantitative methods. On these grounds we will prioritize global qualitative analysis over local quantitative analysis.

(Wills and Pothos, 2012a; 2012b) have convincingly argued that *relative adequacy*, defined in terms of the number and proportion of irreversible, *ordinal* successes, might be a useful metrics for model evaluation and comparison. Central to their approach is the concept of *irreversible success*, that is, success in the absence of arbitrarily variable free parameters. In addition, parameters should be determined at the level of the domain of phenomena that the model is intended to address, not at the level of individual experiments.

This seemingly uncontroversial proposal, that a model that accommodates more successes is, other things being equal, a better model, contrasts sharply with current practice in classical conditioning research, which is to examine in depth the results of a single or a handful of experiments, rather than to seek breadth. Moreover, some researchers insist that model parameters should be derived independently on each occasion. These practices make the evaluation and comparison of computational models of classiscal conditioning harder. To circumvent the difficulties posed by using arbitrary free parameters, (Schmajuk and Alonso, 2012) required the authors to use fixed parameters across all simulations (notice, however, that we didn't penalize the number of parameters à la BIC). However, the fact that most models were tested against small datasets remains an issue. The results in terms of numer of parameters and number of phenomena replicated are shown in Table 2. We are not disputing that the models in (Schmajuk and Alonso, 2012) may account for more results than those explictely reported. However we can only evaluate the models in the light of the evidence provided.

Of course, the meaning of these results is debatable. Nevertheless, it gives researchers in the area a guide of the predictive power of the models. In terms of the number of phenomena replicated, it seems that SLGK is the most comprehensible model. On the other hand, LCT uses only one parameter – which makes us wonder about its real value. It is preferable to endorse models whose verbal description allows some understanding of the

Table 2: Qualitative analysis results.

Model	Number of parameters	Number of phenomena replicated
SLGK	11	82
GP	7	39
AMAN	16	38
SOCR	5	38
TD	11	10
LCT	1	16
PHK+	5	5
MKM/APECS	Unclear	Not fixed

model's processes in psychological terms. This property, that Willis and Pothos call *penetrability* is important, particularly in cases where computational models are taken as psychological models by proxy rather than as formal expressions of psychological models (see, Alonso and Mondragón, 2012).

REFERENCES

Alonso, E., & Mondragón, E. (Eds.)(2011), *Computational Neuroscience for Advancing Artificial Intelligence: Models, Methods and Applications*, Hershey, PA: IGI Global.

Alonso, E., & Mondragón, E. (2012), Uses, Abuses and Misuses of Computational Models in Classical Conditioning. In N. Rußwinkel, U. Drewitz & H. van Rijn (Eds.), *Proceedings 11th International Conference on Cognitive Modeling* (ICCM-12), pp. 96-100. Berlin, Germany: Universitaetsverlag der TU Berlin.

Alonso, E., Mondragón, E., & Fernandez, 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.

Alonso, E., & Schmajuk, N. (2012), Computational Models of Classical Conditioning guest editors' introduction, *Learning & Behavior*, 40(3), 231-240.

Baum, W. M. (1983), Matching, Statistics, and Common Sense, *Journal of the Experimental Analysis of Behavior*, 39, 499- 501.

Dayan, P., & Abbott, L. F. (2001). *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. Cambridge, MA: MIT Press.

Hall, G. (2002), Associative structures in Pavlovian and instrumental conditioning. In H. Pashler, S. Yantis, D. Medin, R. Gallistel an J. Wixted (Eds.), *Stevens' Handbook of Experimental Psychology*, Volume 3, pp 1-45. Hoboken, NJ: John Wiley and Sons.

Haselgrove, M., & Hogarth, L. (2011), *Clinical Applications of Learning Theory*. London, UK: Psychology Press.

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

Mondragón, E., Alonso, E., Fernandez, A., & Gray, J. (2013a), A Rescorla and Wagner simulator with context compounds, *Computer Methods and Programs in Biomedicine*. DOI: 10.1016/j.cmpb.2013.01.016.

Mondragón, E., Gray, J., & Alonso, E. (2013b), A Complete Serial Compound Temporal Difference Simulator for Compound stimuli, Configural cues and Context representation, *Neuroinformatics*. DOI: 10.1007/s12021-012-9172-z.

Pearce, J. M., & Bouton, M. E. (2001), Theories of associative learning in animals. *Annual Review of Psychology*, 52, 111-139.

Schachtman, T. R., & Reilly, S. (2011), *Associative Learning and Conditioning Theory: Human and Non-Human Applications*. Oxford, UK: Oxford University Press.

Schmajuk, N. A. (1997), *Animal Learning and Cognition: A Neural Network Approach*. Cambridge, UK: Cambridge University Press.

Schmajuk, N. A. (2010a), *Mechanisms in Classical Conditioning: A Computational Approach*. Cambridge, UK: Cambridge University Press.

Schmajuk, N. A. (2010b)(Ed.), *Computational Models of Conditioning*. Cambridge, UK: Cambridge University Press.

Schmajuk, N. A., & Alonso, E. (Eds.)(2012). Computational Models of Classical 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.

Schultheis, H., Thorwart, A., & Lachnit, H. (2008b), Rapid-REM: A MATLAB simulator of the replaced elements model, *Behavior Research Methods*, 40, 435-441.

Shanks, D.R. (1995), *The Psychology of Associative Learning*. Cambridge, UK: Cambridge University Press.

Shiffrin, R. M., Lee, M. D., Kim, W., & Wagenmakers, E.-J. (2008), A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods. *Cognitive Science*, 32, 1248–1284.

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.

Vogel, E. H., Castro, M. E., & Saavedra, M. A. (2004), Quantitative models of Pavlovian conditioning, *Brain Research Bulletin*, 63, 173-202.

Wills, A.J. & Pothos, E.M. (2012a), On the adequacy of current empirical evaluations of formal models of categorization, *Psychological Bulletin*, 138, 102-125.

Wills, A. J. & Pothos, E. M. (2012b) On the adequacy of Bayesian evaluations of categorization models: Reply to Vanpaemel & Lee (2012), *Psychological Bulletin*, 138, 1259-1261.