Neurobiology of Learning and Memory 108 (2014) 119–135

Contents lists available at ScienceDirect

Neurobiology of Learning and Memory

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

Review

The error in total error reduction

James E. Witnauer [a], Gonzalo P. Urcelay [b], Ralph R. Miller [c,*]

[a] Department of Psychology, State University of New York at Brockport, USA
[b] Department of Experimental Psychology, University of Cambridge, UK
[c] Department of Psychology, State University of New York at Binghamton, USA

ARTICLE INFO

Article history:
Available online 25 July 2013

Keywords:
Within-compound associations
Cue competition
Associative learning
Conditioned inhibition
Pavlovian conditioning
Total error reduction
Local error reduction

ABSTRACT

Most models of human and animal learning assume that learning is proportional to the discrepancy between a delivered outcome and the outcome predicted by *all* cues present during that trial (i.e., total error across a stimulus compound). This total error reduction (TER) view has been implemented in connectionist and artificial neural network models to describe the conditions under which weights between units change. Electrophysiological work has revealed that the activity of dopamine neurons is correlated with the total error signal in models of reward learning. Similar neural mechanisms presumably support fear conditioning, human contingency learning, and other types of learning. Using a computational modeling approach, we compared several TER models of associative learning to an alternative model that rejects the TER assumption in favor of local error reduction (LER), which assumes that learning about each cue is proportional to the discrepancy between the delivered outcome and the outcome predicted by that specific cue on that trial. The LER model provided a better fit to the reviewed data than the TER models. Given the superiority of the LER model with the present data sets, acceptance of TER should be tempered.

1. Introduction

One measure of the importance of a scientific framework is its influence and explanatory potential across different levels of analysis. In this sense, the associative tradition has been highly successful. Associative ideas have influenced neuroscientific theories (e.g., Schultz, 1998), theories of animal learning (e.g., Rescorla & Wagner, 1972), and connectionist models of cognition (e.g., Rumelhart, Hinton, & Williams, 1986). In parallel with the widespread use of the associative framework, total error reduction (TER) is often used to model changes in the associative structures that support behavioral control and cognitive functioning. The TER view asserts that learning is driven by (and functions to reduce) the difference between predicted and actual events. In recent decades, there have beenseveral reviews of the TER approach to learning (e.g., Gluck & Bower, 1988; Niv & Schoenbaum, 2008; Schultz, 1998). Each of these reviews has presented the TER approach favorably. They often cite neurophysiological correlates of total error signals (e.g., Schultz, Dayan, & Montague, 1997), behavioral tests of TER (e.g., Kamin, 1968), and the widespread use of TER in connectionist modeling (e.g., Nosofsky, Kruschke, & McKinley, 1992). The present review offers a different perspective. We begin by defining TER and reviewing some important applications of TER in behavioral, neuroscientific, and cognitive models. Learning in many influential associative models of Pavlovian conditioning as well as connectionist and neural network models of cognition-seems to be fundamentally similar and based on TER mechanisms, which is also consistent with recordings of neural activity (e.g., Schultz & Dickinson, 2000) and responsiveness to local or systemic pharmacological manipulations (e.g., Lattal & Bernardi, 2007; McNally & Westbrook, 2006). In the present review, a computational modeling approach is used to test TER at the neuroscientific, behavioral, and human cognitive levels of analysis. We demonstrate that a simpler local error reduction (LER) model produces a better fit than several TER models with respect to several important sets of data.

1.1. TER in models of human and nonhuman animal cognition

Early models of classical conditioning (e.g., Bush & Mosteller, 1955) predominantly used LER learning algorithms, which assume that both conditioned stimulus (CS)-unconditioned stimulus (US) contiguity and local error are conjointly necessary for learning. Local error is the difference between the magnitude of the US that is received and the strength of the predicted US with the prediction being based on only the specific CS for which the change in associative strength is being calculated even if other CSs were present on that trial. Such models lacked a mechanism for explaining how

* Corresponding author. Address: Department of Psychology, SUNY-Binghamton, Binghamton, NY 13902-6000, USA. Fax: +1 (607) 777 4890.
E-mail address: rmiller@binghamton.edu (R.R. Miller).

1074-7427/$ - see front matter
http://dx.doi.org/10.1016/j.nlm.2013.07.018

cues interact with each other in associative learning situations. The first application of the TER approach in psychology was in the animal learning literature and was prompted by early experiments in cue interactions (e.g., blocking). Kamin (1968) found that when a target stimulus (X) is paired with the unconditioned stimulus (US) in the presence of a well-established signal (A) for the US, responding to X is reduced relative to a situation in which X is paired with the US in the presence of a previously neutral stimulus. Kamin postulated that so-called blocking reflects a fundamental limitation of simple contiguity theory because X enjoys equal contiguity with the US in the two conditions. Indeed, Kamin argued that surprise upon receiving the US, in addition to X–US contiguity, is necessary for learning about X. Thus, in blocking A should reduce the unexpectedness (i.e., surprisingness) of the US presented during AX–US pairings, and consequently reduce learning about X. The degraded contingency effect is related to blocking and also contributed largely to the subsequent development of TER models (Rescorla, 1968). In degraded contingency situations, unsignalled presentations of the US reduce responding to X when they are interspersed among X–US pairings relative to a control condition in which the unsignalled USs are omitted. Thus by the late sixties, both the degraded contingency effect and conventional (two-phase forward) blocking challenged existing models of learning by suggesting that good contiguity between a cue and its outcome and local error are not sufficient to explain all instances of associative learning.

Based on Kamin's (1968) discovery of blocking and Rescorla's (1968) discovery of the degraded contingency effect, Rescorla and Wagner (1972) developed the first computational model of TER in psychology. According to the Rescorla–Wagner model, contiguity and surprise (captured here as total error across all cues present on a given trial) are necessary conditions for changes in the strength of an association between X and the US. The critical development in the Rescorla–Wagner model was its error term. In this model, learning is directly related to the difference between the maximum associative strength that the US can support (λ) and the total strength of the US expected by the subject based on the sum of the associative strengths of *all* stimuli present at the time of each X–US pairing (ΣV). Changes in associative strength function to reduce total error ($\lambda - \Sigma V$). This model also assumes that the saliences (i.e., associability) of X and the US are important determinants of the rate of associative learning. According to the model, the change in the strength of the X–US association on a given trial is:

$$\Delta V_{\mathrm{X-US}} = \alpha_{\mathrm{X}} * \beta_{\mathrm{US}} * (\lambda_{\mathrm{US}} - \Sigma V_{i-\mathrm{US}}) \quad (1)$$

where α_X is a free parameter representing the salience of X, β_{US} similarly represents the salience of the US (typically two values: one for US present and a second lower one for US absent), λ_{US} represents the maximum associative strength supportable by the experienced US, and $\Sigma V_{i-\mathrm{US}}$ represents the sum (over all cues present) of the strengths of the CS–US associations. This model asserts that the X–US associative strength is updated trial-by-trial using the following equation:

$$V^{(n+1)}_{\mathrm{X-US}} = \Delta V^{(n)}_{\mathrm{X-US}} + V^{(n)}_{\mathrm{X-US}} \quad (2)$$

where n indexes the last trial completed. In other words, the strength of the X–US association at the beginning of a trial is equal to the associative strength before the most recently completed trial plus the change in the associative strength on the most recently completed trial.

The Rescorla–Wagner model explains both the blocking and degraded contingency effects in a straight forward manner. In blocking situations, when X is paired with the US in the presence of a well-established blocking cue (A), total error ($\lambda - \Sigma V_{i-\mathrm{US}}$) is expected to be low relative to a control condition in which a neutral cue replaces the blocking cue because the blocking cue contributes to the expectation of the US, whereas a previously neutral control cue fails to evoke US expectation. Thus, learning about X should be disrupted in the blocking condition because expectation of the US is supported by A. In degraded contingency situations, unsignalled presentations of the US in the training context presumably increase the strength of the context-US association, which then increases US expectation during X–US pairings in the same context and consequently blocks learning about X. The Rescorla–Wagner model has proven to be a powerful model of Pavlovian conditioning that correctly predicts a number of important animal learning phenomena (e.g., Rescorla, 1970, 1971, for reviews, see Miller, Barnet, & Grahame, 1995; Siegel & Allan, 1996). The Rescorla–Wagner model also inspired theoretical and empirical developments across divergent levels of analysis. For example, the Rescorla–Wagner model explains some phenomena within human cognition (e.g., category learning). The observation of blocking in human contingency learning prompted researchers to consider the contribution of total error to learning of contingent relationships by humans (Shanks, 1985). This prompted the development of models of human contingency learning based directly on the Rescorla–Wagner model (e.g., Van Hamme & Wasserman, 1994; detailed below). In the category learning literature, one of the earliest applications of TER was Gluck and Bower's (1988) adaptive network (see also McClelland & Rumelhart, 1985). In their influential paper, Gluck and Bower noted the formal similarity between TER as implemented in the Rescorla–Wagner model and TER in connectionist models of human cognition. While the details of stimulus representation in their adaptive network model differed slightly from that of the Rescorla–Wagner model (largely in order to encompass category learning), the learning rule (TER) was identical. Based on mechanisms similar to those used to explain blocking, this model correctly predicts that participants often neglect baseline rates of categories in judgements of a feature's likelihood of predicting a rare category.

The PDP (Parallel Distributed Processing) research group (Rumelhart et al., 1986) provided important theoretical contributions to cognitive science that were instrumental in the widespread acceptance of connectionist models of cognition. One such contribution was the application of TER to multilayered connectionist models of human cognition. These so-called backpropagation models of cognition are a powerful class of connectionist models that are capable of learning complex, nonlinear (e.g., exclusive-or) input–output mappings. Backpropagation models are characterized by the use of one or more layers of hidden units between input (sensory) and output (response) units. When a type of TER learning algorithm is used, these backpropagation models are capable of learning to solve complex discriminations (e.g., Pearce, 1994). The combination of TER and hidden layers makes these models considerably more powerful and complex than their two-layer connectionist counterparts.

Many backpropagation models are intended to be brain-like in several ways and are therefore sometimes referred to as artificial neural networks. First, units in backpropagation networks are assumed to be neuron-like in terms of their activation functions. The sigmoid activation function that is fundamental to backpropagation models loosely corresponds to the relationship between stimulation and firing rate in real neurons. Second, unlike models of cognition that ordinarily use discrete representation of stimuli (e.g., Gluck & Bower, 1988; Rescorla & Wagner, 1972), stimuli are assumed to be represented by distributed patterns of activation across the hidden layer of the network. This corresponds to the distributed neural representational systems that are presumably used by the brain and rejects the local representational approach. Third, these models assume feedforward processing, meaning that