A Double-Error Correction Computational Model of Learning

Niklas H. B. Kokkola

Supervisor: Dr. Eduardo Alonso
First Supervisor
Dr. Chris Child
Second Supervisor
Dr. Esther Mondragón
External Supervisor

This dissertation is submitted for the degree of

Doctor of Philosophy

City, University of London
Department of Computer Science

October 2017

Table of contents

List of figures

List of tables

Contributions

Conference papers

- XXVI Meeting of the Spanish Society for Comparative Psychology
  September 10-12, 2014, Braga, Portugal
  A Double Error Correction Model of Classical Conditioning.
  N. Kokkola, E. Mondragón, E. Alonso

- XXVII Meeting of the Spanish Society for Comparative Psychology
  September 9-11, 2015, Seville, Spain
  A Double Error Model of Classical Conditioning: Integrating Associatively Mediated Effects.
  N. Kokkola, E. Alonso, E. Mondragón

- Associative Learning Symposium XX
  22-24 March, 2016, Gregynog, UK
  A Double Error-Term Model of Classical Conditioning: Reconciling Associatively Modulated Learning.
  N. Kokkola, E. Alonso, E. Mondragón

- XXVIII Meeting of the Spanish Society for Comparative Psychology
  12-14 September, 2016, Barcelona, Spain
  A Double Error-Correction Model of Classical Conditioning: Dual Stimulus Associability Process.
  N. Kokkola, E. Mondragón, E. Alonso

Journal papers

- E. Mondragón, E. Alonso, and N. Kokkola (2017), Associative Learning Should Go Deep. Trends in Cognitive Science. DOI: http://dx.doi.org/10.1016/j.tics.2017.06.001.

- N. Kokkola, E. Mondragón, E. Alonso, A Real-Time Double Error-Correction Model of Associative Learning (in preparation)

I would like to dedicate this thesis to my father, family, and friends for their support; and to my supervisors for their guidance.

Declaration

I hereby declare that I grant powers of discretion to the University Librarian to allow the thesis to be copied in whole or in part without further reference to the author. I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements.

Niklas H. B. Kokkola
November 2017

Abstract

In this thesis, the Double Error model, a general computational model of real-time learning is presented. It builds upon previous real-time error-correction models and assumes that associative connections form not only between stimuli and reinforcers, but between all types of stimuli in a connectionist network. The stimulus representation uses temporally-distributed elements with memory traces, and a process of expectation-based attentional modulation for both reinforcers and non-reinforcing stimuli is introduced. A modified error-correction learning rule is proposed, which incorporates both an error-term for the predicted and predicting stimulus. The static asymptote of learning familiar from other models of learning is replaced by a similarity measure between the activities of said stimuli, resulting in more temporally correlated stimulus representations forming stronger associative links. Associative retrieval based on previously formed associative links result in the model predicting mediated learning and pre-exposure effects. As a general model of learning, it accounts for phenomena predicted by extant learning models. For instance, its usage of error-correction learning produces a natural account of cue-competition effects such as blocking and overshadowing. Its elemental framework, which incorporates overlapping sets of elements to represent stimuli, leads to it predicting non-linear discriminations including biconditional discriminations and negative patterning. The observation that adding a cue to an excitatory compound stimulus leads to a lower generalization decrement as compared to removing a cue from said compound also follows from this representational assumption. The model further makes a number of unique predictions. The apparent contradiction of mediated learning in backward blocking and mediated conditioning proceeding in opposite directions is predicted through the model's dynamic asymptote. Latent inhibition is accounted for as occurring through both learning and selective attention. The selective attention of the model likewise produces emergent effects when instantiated in the real-time dynamics of the model, predicting that the relatively best predictor of an outcome can sustain the largest amount of attention when compared to poorer predictors of said outcome. The model is evaluated theoretically, through simulations of learning experiments, and mathematically to demonstrate its generality and formal validity. Further, a simplified version of the model is contrasted against other models on a simple artificial classification task, showcasing the

power of the fully-connected nature of the model, as well as its second error term in enabling the model's performance as a classifier. Finally, numerous avenues of future work have been explored. I have completed a proof-of-concept deep recurrent network extension of the model, instantiated with reference to machine learning theory, and applied the second error term of the model to modulating backpropagation in time of a vanilla RNN. Both the former and latter were applied to a natural language processing task.

Chapter 1

Literature Review and Methodology

1.1 Motivation

The motivating problem for the work presented in this thesis was producing a model of associative learning, which could account for both phenomena involving neutral learning and mediated learning. Neutral learning involves associations forming between non-reinforcing (non-reward/punishment) stimuli. Mediated learning in turn involves learning associations between stimuli, one or both of which are not present at a given instant. Both forms of learning are key to building a general model of associative learning and were thus key topics of interest to the present work.

As I sought to create a general, coherent model of associative learning, I further sought to enable the model to predict a wide set of phenomena predicted by extant models of associative learning. The include for instance non-linear discriminations, generalization, attentional effects, as well as cue-competition effects.

To tackle the problem of how a model of learning might account for both mediated and neutral learning, changes and innovations to three integral components of any model of learning were explored. Firstly, the way stimuli are represented. In terms of representa-

tion, a fully-connected network of elements representing stimuli was used. As connections between neutral cues were thereby supported, this seemed like a natural approach to producing accounts of neutral learning. A second common component of a learning model is a process whereby the selective attention (i.e. speed of learning) paid to a given cue is changed as a factor of learning and time. In the case of the present model, I chose to utilize an attentional process based on time-averaged uncertainty, as it seemed to most naturally account for evidence of attentional processing during exposure of a stimulus in the absence of reinforcement. Thirdly, any associative learning model must incorporate a rule for how links between stimuli change as a factor of learning. Here, I chose to explore how a unique asymptote of learning and second error-term could be used to modify a traditional delta-rule, as these two additions seemed like a natural approach to both accounting for the learning component of latent inhibition, as well as how cues learn differentially when they are directly present or retrieved through memories.

1.2 Contributions

The contributions of the introduced DE model, detailed in this thesis, to the fields of learning theory and machine learning have been as follows:

- The DE model introduces a unique second, predictor error-term. This error-term attenuates the speed of learning between a predictor and outcome according to the extent to which the predictor is predicted by other cues and itself. This mechanism is integral to the model's ability to account for various pre-exposure effects. In Chapter 4 we have demonstrated the connection of this second error-term to covariances estimates, and in Chapter 5 we have shown that its use as a simple classifier.

- The model introduces a unique Hebbian-inspired dynamic asymptote of learning, which measures the similarity of activation levels between two elements. Thus, elements with more similar activity levels are capable of forming stronger associations with one another. This in a sense measures the likelihood that the two elements are causally linked, and endows the model with the ability to predict a wide range of learning effects, including mediated learning. Of specific note is that it can account for both backward blocking and mediated conditioning/SPC simultaneously.

- The model introduces revaluation alphas, which track a time-window of uncertainty in the occurrence of reinforced and non-reinforced stimuli. In addition, persistent uncertainty is turned into a source of information in and of itself. This is done by measuring whether the moving-average of stimulus uncertainty is over a threshold, and if so reducing attention to cues of which the occurrence is inherently random. These alphas allow the model to account for the sigmoidal shape of latent inhibition, in addition to numerous other effects.

The DE model similarly, as detailed in Chapter 2 and 3, produces a variety of distinct predictions in the field of learning theory. Some are novel to DE; others, although present to various degrees in other models, arise naturally in DE as a coherent corpus.

- For CSs of equal salience and sufficiently long duration, the best predictor of an outcome will capture the highest amount of attention.

- Associations form between any two stimuli, of which the representations are concurrently active.

- Learned irrelevance is mediated through CS and context-dependent activation of the US representation. Hence, stronger context-US links will produce a stronger effect.

- The sigmoidal shape of latent inhibition is proportional to the attenuation of selective attention paid to the CS, and hence proportional to the length of CS exposure.
- The context specificity of latent inhibition is produced by context to CS associations forming during pre-exposure, and is attenuated by exposing the context in isolation after CS exposure.
- The Hall-Pearce effect arises in part due to the weak-shock trials reducing the selective attention paid to the CS, as well as due to the formation of context-CS associations. As such, conducting the second phase of the treatment in a novel context should significantly attenuate the observed effect.
- The difference in perceptual learning between intermixed and blocked presentations of cues arises in part due to CS-CS association between the reinforced cue and its intermixed associate in the first phase being weaker than the CS-CS association between the reinforced cue and the CS presented in the block of trials. This leads to lower mediated conditioning during the subsequent reinforced trials in the former condition. Hence, preceding the reinforced trials with non-reinforced trials of the subsequently reinforced CS should lead to a smaller difference in perceptual learning between the intermixed and blocked conditions.
- The DE model predicts that a CS, which was exposed in compound with another CS, experiences more latent inhibition than a CS exposed in isolation due to the former condition leading to the subsequently reinforced CS retrieving a representation of its associate. This retrieved associate in return produces a prediction for the retriever, reducing the retriever's speed of association toward the reinforcer. Hence, a further prediction is that should the retriever CS undergo further non-reinforced exposure after its exposure with its associate, the increased latent inhibition effect should be attenuated.

- The DE model predicts that mediated learning effects as a whole can be accounted for as the result of the combination of associative retrieval through within-compound associations and learning proceeding according to the similarity in activation between the representations of two stimuli (whether directly present or retrieved).

- Further, the DE model predicts that the apparent contradiction in learning proceeding in opposite directions in backward blocking and mediated conditioning/SPC procedures arises due to the prior training endowing the retrieved cue with respectively moderate or no associative strength. Hence, since the asymptote of learning is respectively lower and higher during the mediated learning, extinction and acquisition is observed.

- As the temporal overlap between cues influences the asymptote of learning for associations forming between them, the model additionally predicts that backward blocking is respectively facilitated and attenuated by more phase 1 training according to whether the CS used is of long or short duration.

- The model predicts that in general mediated conditioning, SPC, and BSP will occur in proportion to the extent to which the retrieved cue is retrieved. As such, more Phase 1 training should strengthen the observed effect.

- The DE model predicts the configural discrimination in Chapter 3 through the context becoming highly excitatory. Therefore, this hypothesis is falsifiable by measuring responding in a further context test phase.

- In terms of timing, the model predicts that the early portions of a sufficiently long CS will tend to become inhibitory toward the US after sufficient training during an acquisition procedure.

1.3 Overview of Work

We initially conducted a review of the existing models in the animal learning literature. Based upon this review, we have developed a real-time model of learning (the "Double Error" or "DE" model henceforth), which extends concepts of multiple prominent models of learning in the field. These models and the learning phenomena they aim to explain are detailed in Section 1.2, of the literature review. The crux of this new model is that memories formed through associative learning interact with one another in a biologically plausible and mathematically concise way (elaborated in Chapter 2), instantiating a basic form of causal reasoning. The model can explain neuro-psychological phenomena that existing theories of learning struggle with, while maintaining predictive generality and parsimony.

We have conducted a mathematical analysis of the model's relation to similar algorithms and theories in both the psychological learning (Chapter 1) and machine learning fields (Chapter 5), demonstrating that it subsumes some of the predictions and functioning of models in these areas, while allowing for novel behaviour not demonstrated by extant models of learning. The mathematical convergence properties of the DE model are explored in Chapter 4.

We have implemented the model in open-source specialist-user friendly Java simulators, which can be run on most operating systems. This allows for us to explore the inner workings of the model, as well as to produce empirical predictions to be compared against data obtained from our own data-sets as well as from published experiments. It is of considerable utility, as the complex non-linear interactions of the model preclude results from being understood a priori, except in very general cases. The results of the simulations can be saved to a design-file, can be used to produce and save various interactive graphs, and exported to an excel format.

In developing the DE model, we have explored numerous variations of its core principles and their results on the model's operation. We have worked with different mathematical representations of stimuli (how many and what kind of components represent a memory), explored different rules of how precisely they interact with one another in terms of learning, analysed how commonalities between events in the world are taken into account in learning causality, as well as tried different variations of how the contextual environment within which learning takes place should be modelled.

We have had success in these endeavours. We have both accounted for classical learning phenomena in the literature that other models can account for, while in many cases explaining further phenomena which many models struggle explaining. For those phenomena that alternative models accommodate, our model provides a distinct, novel explanation. These results have been presented in various conferences and published in academic journals. These research outputs are presented in Chapter 3 of this thesis.

Further, we have evaluated the performance of a simplified version of the Double Error model in a simple contingency learning/classification task in Chapter 5, demonstrating performance on-par to a naive Bayes and single-layer perceptron model.

Finally, multiple avenues of future work have been envisioned. For instance, I have begun experimentation on application of the DE model to NLP tasks in two separate ways. The first uses a 'deep' instantiation of the Double Error model as a recurrent network based on the principle of predictive coding, while the second uses an open-source recurrent neural network (RNN) implementation that is modified to include a predictor error-term to modulate the back-propagation in time of the error-gradient. These experiments are detailed in the

future work section of the conclusion.

There have been various obstacles in conducting our research, but they have fortunately been overcome. The nature of a highly non-linear model is that it cannot be studied analytically, and therefore its workings are only fully understood through the results of simulating it. Thus, it has been difficult at times to ascertain whether some erroneous result is the product of a flaw in the exact formulation of the model, whether it is the result of a bug in the implementation of the model, or possibly another model-independent bug. We have devised various successful approaches to dealing with such difficulties, and I have received considerable guidance on where exactly to look when obstacles have occurred. In a similar vein, the simulations we run require the execution of very computationally expensive calculations, as well as the dynamical storage of large amounts of RAM data due to data-arrays with millions of entries. We have therefore, to avoid the simulator from running out of computational resources, had to commit time to reducing memory overheads in the code, saving and retrieving data to and from the hard drive dynamically, as well as used other workarounds to smooth out issues with running highly complex simulations on desktop computers.

1.4 Summary of the Model

We will overview the overall processes of the DE model on a given trial in this section. The detailed equations for the components of the model are introduced in Chapter 2. The pseudo-code for all these processes are displayed in Section 4.1 of this thesis.

The DE model instantiates a connectionist network consisting of elements (nodes), which belong to individual stimuli or are shared in common between pairs of stimuli. At each time-point within a trial or ITI period, the model calculates the direct (sensory) activation of an element according to a semi-Gaussian function with a specified mean (determined by which time-wave the element is sampled from). Next, the associative activation of elements is calculated based on the predictions generated for said element on the previous time-point by all other elements. The overall activation of an element is then taken to be which ever is larger: the direct activation or the discounted associative activation.

The model then calculates the revaluation alphas individually for each element. This occurs in proportion to the activity of the element (such that more active elements experience faster changes in their alpha values). The value that the alphas change towards is the time-averaged mean error of all reinforced (α_r) or non-reinforced (α_n) elements. If this time-averaged mean error value crosses a threshold, the model decays the respective alpha value at each time-point on which this condition remains true.

Finally, the DE model calculates the learning between each pair of elements (in both directions). First, the asymptote of learning is calculated. This asymptote is higher for two elements, which have more similar overall levels of activation. This dynamic asymptote enters into the error-term of the predicted element, along with the summed predictions from all other elements for this predicted element. Similarly, the predictor error-term is calculated

as the discrepancy between the predictor element's overall activity and predictions made for it by other elements.

The weight from the predictor element to the predicted element is then calculated as the product of the two error-terms mentioned (with the absolute value of the predictor error-term being taken), along with the saliences of the two elements, the overall activations of the two elements, and the revaluation alpha from the predictor to the outcome (i.e. α_r or α_n if the outcome is respectively a reinforcer or a non-reinforcer).

Although the DE model may seem formally complex, its essence is rather simple. Elements of the CS, context, and US are treated for the most part equivalently in that their learning of associations at each time-point is calculated using the same equation. The overall activation of an element is simply a factor of its sensory input and predictions made for it by other elements. The revaluation alphas of a given element track the overall uncertainty of reinforced and non-reinforced events in a straightforward manner. Finally, the direction of learning in the model between two elements is dictated by the activation similarity of the two elements along with cue-competition; with other factors simply modulating the magnitude of this learning.

1.5 Literature Review

List of abbreviations for Chapters 1-3

$-$	a non-reinforced trial
$+$	a reinforced outcome
α:	learning rate/associability/salience
A1:	A1 SOP activation state
A2:	A2 SOP activation state
BB:	Backward blocking effect
BSP:	Backward bensory preconditioning
CR:	conditioned response
CS:	a conditioned (motivationally neutral) stimulus
DE:	Double Error model
FSP:	Forward sensory preconditioning
ISI:	Inter-stimulus interval
ITI:	Inter-trial interval
LI:	Latent inhibition effect
LTM:	Long term memory
MSE:	Mean-squared error
PH:	Pearce-Hall model
REM:	Replaced Elements model
RR:	Retrospective revaluation
RW:	Rescorla-Wagner model
SLGK:	SLGK model
SOP:	SOP model
STM:	Short term memory
TD:	Temporal Difference model
UR:	unconditioned response
US:	an unconditioned stimulus (reward)
V:	Associative strength of a link

Associative Learning

In classical conditioning, the repeated co-occurrence of two events (e.g. an odour or tone), S1 and S2 is assumed to result in an association between the internal representations of these stimuli, which entails that the presence of S1 (the CS or 'predictor') will come to activate the internal representation of a S2 (the predicted stimulus or outcome from now on) in the brain of an animal. When the outcome is a biologically relevant (e.g. food) stimulus (unconditioned stimulus, US) able to elicit an unconditional response (UR) by itself, learning between the CS and US results in the acquisition of a new pattern of behaviour: the sole presence of the CS engenders a conditioned response (CR) similar to the UR (Figure 1.1). The response elicited by the CS is then assumed to express the strength of the association between the CS and the outcome (Pavlov, 1927), and therefore indicates that the outcome is anticipated or predicted by the CS. Associative learning phenomena have been replicated across numerous species (Nader, 2003), the neural correlates of learning have been extensively studied (Gomez et al., 2001; Kobayashi and Poo, 2004; Panayi and Killcross, 2014), and its evolutionary origins are beginning to be elucidated (Ginsburg and Jablonka, 2010).

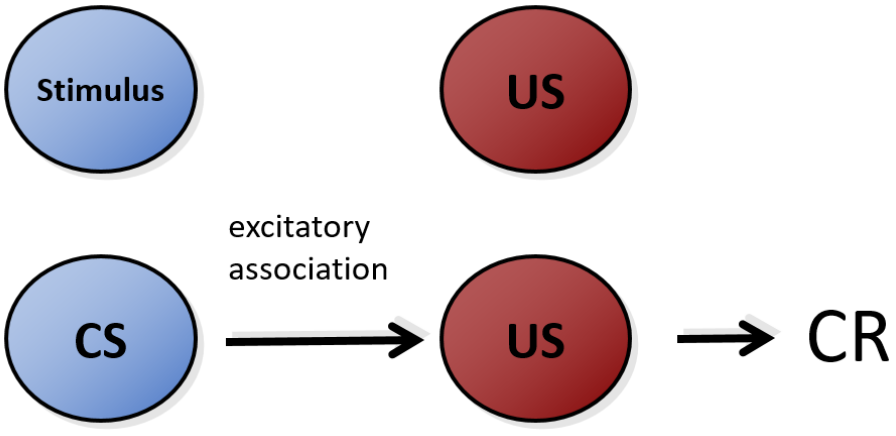


Fig. 1.1 Conditioning. The pairing of a stimulus with a US engenders the formation of an excitatory association from the (now) conditioned stimulus to the US. This association invokes the production of a conditioned response (CR).

The acquisition of a CR, and thus of the $CS \rightarrow US$ link, follows a negatively accelerating, monotonically increasing curve over trials (but not always, see (Gallistel et al., 2004; Glautier, 2013)) . The associative link's rate of change is thereby proportional to the discrepancy between the expectation and presence of the outcome, i.e. the prediction error. This relation was initially mathematically formalised in linear-operator models such as Hull's early quantitative theory of learning (Hull, 1943) and stochastic theories of conditioning (Blough, 1975; Bush and Mosteller, 1955). A secular trend in modelling since then has been the expansion of the ontology of internal stimulus representations and learning processes governing the formation of associations between said representations. This increase in model complexity has expanded the quantity of phenomena that can be accounted for by models of learning (Alonso et al., 2014; Alonso and Schmajuk, 2012; Balkenius and Morén, 1998; Pearce and Bouton, 2001), and has been propelled and refined by experimental data acting as an arbiter of these models. For instance, evidence for the hypothesis that cues compete with one another for associative strength necessitated advancements from linear-operator error-terms. As an example, the phenomenon of blocking (Amundson and Miller, 2008; Jennings and Kirkpatrick, 2006; Kamin, 1968, 1969; Kohler and Ayres, 1979) showed that when the acquisition training for a cue A is followed by acquisition training with a compound of two cues, AB, the novel cue B acquires next to no conditioning. Thus, it seems as if the associative link formed by cue A to the outcome prevents the subsequent formation of an equivalent link between B and the outcome. This result indicates that the processing of the outcome plays a significant role in determining the maximal amount of learning supported between itself and the CS.

The Rescorla-Wagner Model

Such cue-competition during learning is formalized most prominently by the Rescorla-Wagner (RW) model with its 'global' prediction error-term (Rescorla and Wagner, 1972) that incorporates within it a sum of all associative links to the US. This is an innovation upon other linear-operator error-terms, as for instance the Hull error-term in contrast only incorporates the associative strength of an individual CS. Hence, the learning between a CS and US in the RW model is driven by the total discrepancy between the US presence at a given trial and the expectation elicited for it by all cues:

$$\Delta V_i = \alpha_i \beta_j x_i (z_j \lambda_j - \sum x_i V_i) \tag{1.1}$$

The activation of the CS and US representation, corresponding to the presence of the CS and US, are denoted by $0 \leq x_i \leq 1$ and $0 \leq z_j \leq 1$. The $0 \leq \alpha_i \leq 1$ and $0 \leq \beta_j \leq 1$ terms are respectively saliences of the CS_i and US_j, representing stimulus intensity. They modulate learning by increasing conditioning speed for stimuli which are psychophysically or otherwise important for an animal. Thus, a loud sound is assumed to have a higher salience than a faint one, and likewise for a food reward which an animal is known to prefer. V_i denotes the strength of the associative connection between the adaptive unit of stimulus 'i' and the US. Each stimulus competes with all other stimuli present on a trial for a fixed amount of reinforcement by the US. This was achieved through the incorporation in the error-term of the total associative strength of all stimuli present on a given trial, taken to denote the limited capacity of reinforcement available for stimuli to compete over. Importantly, it was also assumed that the total associative strength on a trial was a linear summation of individual associative strengths of stimuli to the US.

The total associative strength is calculated in the model using the following equation:

$$V^T = z_j \lambda_j - \sum x_i V_i \tag{1.2}$$

Here λ is the asymptote of learning (the maximal link strength supported), z_j and x_i denote the presence of the outcome j and stimulus i respectively (and therefore the activation of their internal representation), and V_i denotes the strength of the associative connection between the adaptive units of the stimulus 'i' and the US j. Thus, a given stimulus is only responsible for eliciting a partial expectation of the outcome. The size of this part is influenced by prior learning other stimuli have undergone, as well as their saliences.

The learning rule of the model updates the previous associative strength value between a stimulus i and US j by incrementing it by ΔV_i:

$$V_i^t = V_i^{t-1} + \Delta V_i^t \tag{1.3}$$

Predictions of the Rescorla-Wagner Model

The RW model's competition for fixed reinforcement is an elaboration upon earlier error-correction models. Previous *linear sampling* theories assumed an independence from the influence of other CSs in the learning of a CS-US relation. Significantly, by rejecting this notion, the RW model could explain phenomena such as blocking discovered by Kamin (Kamin, 1969). The RW model could account for the loss of associability of the blocked stimulus by positing that a combined associative strength value of all stimuli present in a trial is involved in modulating the size of the error term. Therefore the prior conditioning between blocking stimulus A and the outcome leads to a smaller error-term and thus learning between

B and the US. The aggregation of all associative strength values in a trial similarly allowed the RW model to account for the effect of overshadowing (Mackintosh, 1976), wherein a weak conditioned stimulus accrues less conditioning when presented together with a more salient stimulus. RW further correctly predicts a number of previously unknown phenomena such as overexpectation (Lattal and Nakajima, 1998), superconditioning (Dickinson, 1977), and protection from extinction (Rescorla, 2003). Overexpectation occurs through a procedure in which two independent CSs have been conditioned to an asymptotic level. Subsequently presenting these two CSs as a compound produces a prediction for the outcome, which exceeds its capacity for reinforcement; thereby producing inhibitory learning. Superconditioning is produced when a predictor of an outcome is presented together with a novel CS, turning the latter into a conditioned inhibitor of the outcome. This conditioned inhibitor is thence paired with another novel CS in conjunction with the reinforced outcome, which results in this second novel CS gaining excitatory associative strength more rapidly than if it were trained in isolation with the US. Protection from extinction in return is observed when non-reinforced presentations of a CS together with a conditioned inhibitor result in only a minor loss of associative strength of the CS. Due to these advantages, and despite its shortcomings (see (Miller et al., 1995)), RW is still considered one of the most influential models of classical conditioning.

The successes and failures of the model in accounting for phenomena are recorded in Table 1.1 below, based on the assessment of the model by (Miller et al., 1995, pp. 365-378). Though the number of recorded failures eclipses the number of successes, this does not imply the futility of the model, as "[...] many of the following behavioural phenomena would not have come to light without the model suggesting critical experiments. In the heuristic sense, even its failures (Table 1.2) reflect well on the model." (Miller et al., 1995, p. 369)

Phenomena
Acquisition: response curves negatively accelerating and monotonically increasing
Extinction: response curves negatively accelerating and monotonically decreasing
Stimulus generalization: responding elicited proportional to similarity of stimuli
Discrimination: differential responding to different combinations of stimuli
Tests for conditioned inhibition (i.e. negative association), e.g. summation
Procedures for producing conditioned inhibition (i.e. negative association), e.g. intermixed A+/AB- trials
Patterning (AB+/A-/B- or AB-/A+/B+)
Overshadowing, i.e. a more salient stimulus hinders learning between another stimulus and the outcome
Overexpectation resulting from two excitors: separately reinforcing two stimuli results in supra-maximal responding when they are thence presented in compound
Blocking: A+/AB+ results in less B to outcome learning than AB+ by itself, due to A having acquired associative strength
Unblocking with an increased US: blocking effect decreases when the US intensity is increased in the second phase
Blocking with a reduced US: blocking effect is strengthened when the US intensity is decreased in the second phase of a blocking experiment
Relative validity of cues: associative strength is proportional to the probability of reinforcement
Superconditioning: training a novel cue with an inhibitory cue results in more excitatory learning to the novel cue
US-Preexposure effect: presenting the US by itself prior to training results in weaker learning between a CS and US.
Contingency Effect: Presenting the US by itself between trials results in less excitatory CS-US learning
Trial spacing effect: more temporal separation of trials leads to more excitatory CS-US learning
Instrumental behaviour (association between behaviour and reward) can be explained by the model

Table 1.1 Successes of the RW model in predicting learning effects (Miller et al., 1995).

Phenomena
Spontaneous recovery: responding can resume after extensive extinction training
External disinhibition: presenting a salient yet novel cue before an extinguished cue can produce resumed responding
Reminder-induced recovery from extinction: presenting cues from previous training can produce resumed responding in an extinguished cue
Facilitated and retarded reacquisition after extinction: learning after an extinction treatment can occur at a different pace than before extinction training
Presenting a conditioned inhibitor by itself does extinguish its negative association
Presenting a novel non-reinforced cue with a conditioned inhibitor does not lead to the former acquiring excitatory strength
Non-exclusiveness of conditioned excitation and conditioned inhibition: evidence suggests a cue can be both excitatory and inhibitory toward the same cue depending on the temporal relation between cues
Dependence of negative summation on the degree to which the transfer excitor is excitatory: a conditioned inhibitor can increase responding if the cue it is paired with was previously reinforced with a less intense reward than the one the conditioned inhibitor was paired with
CS pre-exposure effect / Latent inhibition: presenting a CS by itself attenuates subsequent CS-US learning
Potentiation: presenting a salient CS with a less salient CS can sometimes produce more excitatory learning in the less salient CS, contradicting overshadowing
Second-order conditioning: pairing a cue A with an outcome, then cue B with cue A, results in cue B producing responding
Cue-to-consequence effects: CSs and USs from congruent sensory modalities are preferentially associated, e.g. flavours and gastric upset
Dependence of asymptotic responding on CS intensity: more intense CSs produce higher asymptotic responding
Learned irrelevance: uncorrelated CS-US presentations attenuate subsequent correlated CS-US learning with the same cues
One-trial overshadowing: overshadowing can occur within the first trial of training
Overshadowing, blocking, and the US-preexposure effect are reversible.
Extinguishing the context can weaken the inhibitory strength of a conditioned inhibitor
Superconditioning is sometimes not observed
Training cues A and B independently as excitors and then presenting them with novel cue X does not seem to make X a conditioned inhibitor as predicted
Large numbers of compound trials can eliminate overshadowing and blocking deficits
Unblocking can be produced by omission of a second, expected US
Presenting a blocked CS without its blocking associate CS still leads to less excitatory learning than expected
A less salient CS is sometimes unable to block a more salient CS
Evidence exists that associative values change within a trial, i.e. in real-time

Table 1.2 Failures of the RW model in predicting learning effects (Miller et al., 1995).

Rescorla-Wagner and the Summation Assumption

When two or more stimuli are presented together, this presentation is termed a *compound stimulus*. Elemental models, which support strict summation, claim that an animal's response to a compound stimulus will be equal to the sum of the responses, which each individual stimulus would have elicited when presented alone. This response is assumed to reflect the associative strength between the individual stimuli and the outcome. Thus, for instance when two stimuli (A & B) are individually reinforced and come to elicit respectively a CR_A and CR_B, their presentation together should elicit a CR of the magnitude $CR_A + CR_B$. Other interpretations however violate the summation rule, and are supported by experimental evidence in many cases (Razran, 1939). They allow for the associative strength of a compound to be higher or lower than the direct summation of its constituent stimuli. This implies that additional information is encoded by the whole configuration of stimuli, distinct from their components. *Configural* models of learning posit that strict summation of associative strengths does not occur, and in fact animals represent compound stimuli through compound nodes with their own associative strengths. Thus the compound AB has an independent associative strength towards the outcome compared to its constituent stimuli, A and B. Existing elemental models lacked this representational feature and therefore needed a method to render non-linear discriminations. The introduction of *configural cues* (Wagner and Rescorla, 1972) allowed the RW model to account for the failure of summation to materialize in certain compound trials, thereby retaining the structure of an elemental model while explaining the ability of animals to solve non-linear discriminations such as negative patterning (Rescorla et al., 1985) and biconditional discriminations (Rescorla, 1972; Saavedra, 1975). Negative patterning consists of nonreinforced stimulus compound trials being intermixed with reinforced presentations of the individual stimuli. Biconditional discrimination is a procedure of the form $AC+/AD-/BC-/BD+$, where each cue predicts both the occurrence and lack of occur-

rence of reinforcement. Thus the non-linear discrimination can only be solved by attending to the whole compound. These discriminations could not be resolved under the previous elemental frameworks[1]. Configural cues are assumed to emerge when a compound of two or more stimuli is presented (*i.e.* their stimuli representations are active concurrently). This cue acts as a stimulus in the learning process, and competes with other present stimuli. Take for example the case of negative patterning with intermixed A+/B+/AB- trials. The animal will learn to withhold its response during the presentation of the AB compound, as the configural cue *C* emerging from this configuration undergoes inhibitory conditioning towards the US.

The RW model not only accounts for experimentally confirmed effects, but its formulation implied the existence of learning effects not predicted by earlier models or assumed to exist by pre-existing theory. Though the RW model has been successful in this regard (Miller et al., 1995), a summed error term is not the only means of predicting cue-competition results. Models relying on CS-US specific 'local', non-competitive error-terms are capable of accounting for some of these effects by assuming other processes of competition, for instance on attentional competition between the predictors, as postulated by Nicholas Mackintosh (Mackintosh, 1975). Alternatively, the comparator hypothesis model (Miller and Matzel, 1988; Miller and Witnauer, 2016) explains cue-competition results as a retrieval effect. It assumes that when a previously reinforced CS is presented (the target), it re-activates both representations of other CSs paired with the US and the US representation itself. The response elicited by the target CS is then proportional to the degree to which it predicts the US relative to the associative strength of these comparator CSs. Hence cue-competition arises from the interference of other CS-US associations upon the target-US association (Figure 1.2).

[1]Though solutions distinct from configural cues exist to resolve this problem in an elemental fashion. E.g. see (Harris, 2006).

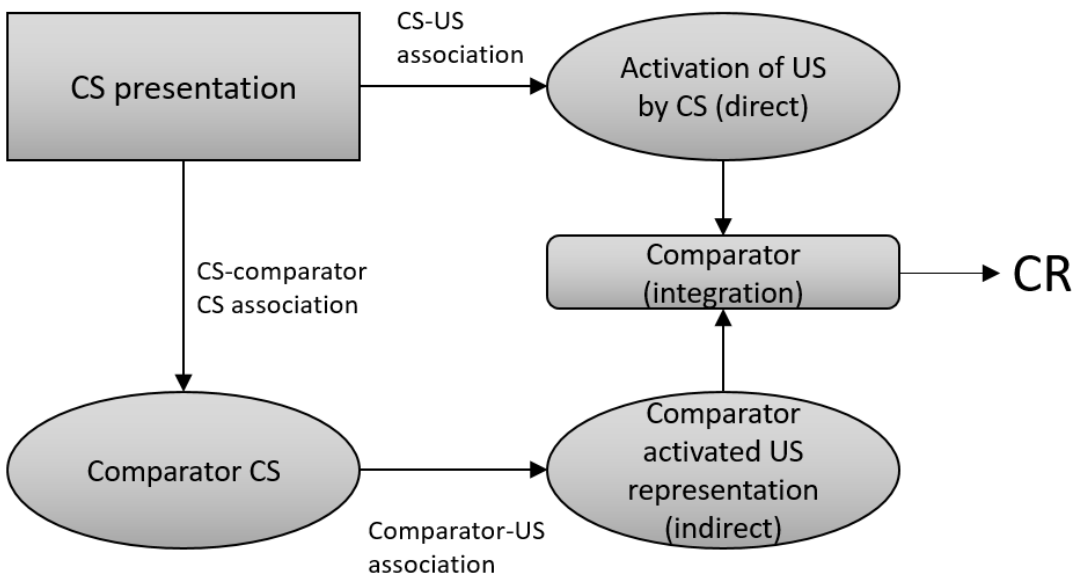


Fig. 1.2 The comparator hypothesis of Miller & Matzel (1988) postulates that the presentation of a CS invokes both a direct US representation, and an indirect activation of the US representation through comparators. These two pathways compete with one another to determine the net CR produced by the CS.

Temporal Difference learning

Since early error-correction models operated based on calculating associative strength on a trial-by-trial basis (i.e. do not model responding and learning within a trial), the realization of real-time learning models extending the trial-based RW model (and other error-correction models) allows for the modelling of time-dependent aspects of learning and temporal relations between stimuli within a given trial. The Sutton & Barto (SB) model (Sutton and Barto, 1981), extended the delta rule used in RW by postulating that the variation in the inputs to a node representing a US was the driver of learning. Hence, both changes in the US sensory input (reinforcement) and changes in the CS contribution to this node (temporal difference) influence the direction of learning. In addition to phenomena predicted by RW, SB accounts for inter-stimulus interval (ISI) effects and an anticipatory CR build-up before the occurrence of a reinforcer (Balkenius and Morén, 1998). It nevertheless produces a

few erroneous predictions, such as that a co-occurring CS and US should become highly inhibitory towards one another. Flaws of the SB model are rectified in the Temporal Difference (TD) model (Sutton and Barto, 1987) through the variations in the CS signal to the US node being dissociated from the US signal itself, and through the introduction of a time-discount γ to signal the uncertainty of future predictions. The TD model uniquely predicts that learning is being driven by the need of the animal to minimise a time-discounted aggregate expectation of future reinforcement. At each time-step, the components of a given CS produce a prediction for the moment-by-moment change in US activation at the next time-step, termed the temporal difference. The difference between this prediction and the actual US activation level results in an error-term (similar to that in the RW model), the prediction error, for that time-step. The equation for calculating this error is as follows:

$$\delta^t = (\lambda^t - (\sum_i x_i^{t-1} V_i^t - \gamma \sum_i x_i^t V_i^t)) \tag{1.4}$$

Here λ^t is the US activation at time t, x_i^t is the activation of a given CS, and V_i is the strength of the associative connection between CS i and the US. The term $\sum x_i^{t-1} V_i^t$ denotes the prediction for the US produced through the summation of associative activations of the US (using current associative links, but CS activation levels from the previous time-step), and the term $\gamma \sum x_i^t V_i^t$ is the equivalent for the current time-step. The current time-step prediction is multiplied by the mentioned discount parameter γ to reflect that the future is always slightly uncertain and therefore more recent stimulus activation carries more weight in calculating errors. The difference between these two prediction terms produces an estimate of what the activation of the US is during the current time-step. The discrepancy between the TD error and the actual US presence produces the overall prediction error for the US.

The error-term is used to adjust associative links between stimuli according to the equation:

$$V_i^t = V_i^{t-1} + \alpha\delta^t x_i^t e_i^t \tag{1.5}$$

Here $0 \leq \alpha \leq 1$ is the salience of the CS, and $0 \leq e_i^t \leq 1$ is an eligibility trace of stimulus i, which modulates the amount of learning a stimulus should receive based on the distance of the stimulus' activation from that of the US. Eligibility traces can either accumulate to values higher than 1 over the duration of a stimulus, or build towards a value of 1 and then continuously undergo replacement by a value of 1[2]. The equation for the latter, the *replacement* eligibility trace, is:

$$e_i^t = min(1, \rho\gamma e_i^{t-1} + x_i^{t-1}) \tag{1.6}$$

The trace therefore is set to and kept at 1 when a stimulus is active, and decays according to $(0 \leq \rho \leq 1) \cdot (0 \leq \gamma \leq 1)$ every time-step when it is not active. This results in stimuli receiving less reinforcement if there is a large gap between when they are active and when the US occurs.

The error-term can produce excitatory or inhibitory learning depending on whether the change in associative strength, ΔV_i^t, is respectively positive or negative. Therefore, the speed of learning between the CS and the US is inversely proportional to the certainty with which the animal can predict the occurrence of the US. This means that when the error of predicting

[2]Accumulating eligibility traces are nevertheless not widely used due to convergence problems produced by values higher than 1.

the US is reduced to zero, learning reaches an asymptote and effectively stops.

As a real-time rendition of the US-processing error-correction learning found in RW, TD can account for the same phenomena as the former on a trial level. In terms of predictive power, the instantiation of the TD model in simulators (Mondragón et al., 2013a; Mondragón, E., Alonso, E., Fernández, A., and Gray, 2012) has produced accounts of higher-order conditioning (Figure 1.3), wherein reinforcement is observed in the absence of a US, through the reinforcement effects of the temporal-difference term itself. It accounts for temporal primacy effects (e.g. (Kehoe et al., 1987)), wherein the presentation of a reinforced serial-compound of the form $CS_A \rightarrow CS_B \rightarrow US$ results in a deficit in CR acquisition to CS_B). It additionally reproduces the retardation effects of ISIs on the level of asymptotic learning (and by extension, CR over time) in an emergent fashion (Balkenius and Morén, 1998) through higher-order conditioning resulting from the temporal-difference term back-propagating to earlier time-points. As the core component of the TD model is its distinct error-term, variations and extensions of TD with differing stimulus representations have been proposed. A complete serial compound model (Moore et al., 1998) postulated that a CS is represented in time by a series of distinct units, each of which becomes active following the previous, in a cascading pattern. This extension produced a response curve in time with closer correspondence to empirical fact.

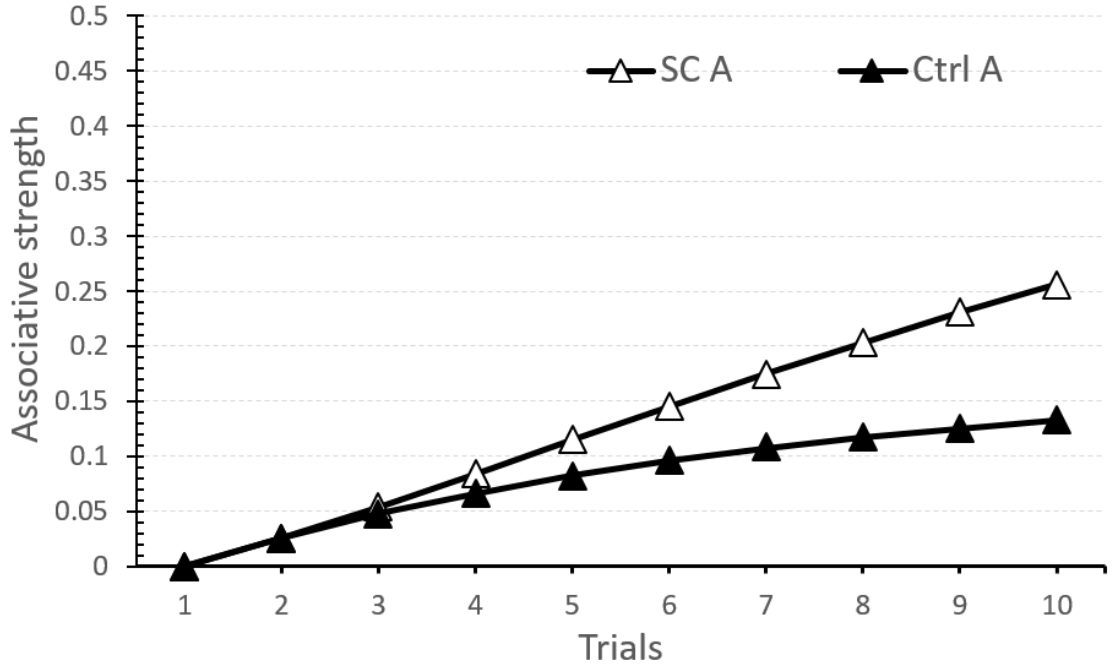


Fig. 1.3 Simulation of the TD model using the TD simulator (Gray, J., Alonso, E., Mondragón, E., and Fernández, 2012). Displayed is the result of simulating acquisition for a reinforced serial compound $A \rightarrow B \rightarrow +$ (Group SC) compared to $A \rightarrow +$ (Group Ctrl). Group SC displays higher-order conditioning.

Work in timing theory as well as hippocampal data inspired the micro-stimulus representation of TD (Ludvig et al., 2009), which instead assumes that the presence of a CS produces a cascade of units to become active in a bell-shaped, Gaussian form, with units that peak later having a correspondingly lower amplitude and higher variance. The advantage of such a stimulus representation is that it reproduces both differential responding early and late during a CS presentation, allows for effective trace conditioning (conditioning with a delay between the CS offset and US onset) due to the persistence of later micro-stimuli, and produces generalization of learning in time due to the significant activation overlap of said micro-stimuli. Extensions of the classical TD formulation have also been motivated by its lack of compound stimuli and configurations in its stimulus representation. Formulations of TD such as a version of CSC (Mondragón et al., 2013b) and SSCC (Mondragón et al., 2014) offer rules for forming such compounds both when stimuli overlap temporally, and when

they are presented in succession. Steady evidence has mounted that predictions errors akin to the TD error are correlated to mid-brain dopamine function (Ludvig et al., 2011; Montague et al., 1996; Niv, 2009; Niv et al., 2012; Schultz, 2004, 2006, 2010; Schultz et al., 1997), specifically with increased dopaminergic activity corresponding to positive prediction errors, and lulls in dopaminergic activity corresponding to negative prediction errors. Furthermore, its efficacy in the domain of machine learning has been strengthened both due to its proven performance in practice, as well as mathematical analysis guaranteeing strong convergence properties toward an optimal policy (Sutton and Barto, 1998).

In conclusion, although alternative approaches to summed error-correction such as the comparator hypothesis model exist and are capable of accounting for crucial cue-competition effects, the fit of the RW and TD error-terms to neuroscientific data correlating prediction errors and dopamine signalling as well as a consideration for parsimony motivate our use of this learning framework of real-time error-correction in the Double Error model we are introducing.

Elemental vs. Configural models: the nature of the stimulus representation

Causal relations in the world are mostly more complex than the linear CS-US relationship learned in a standard acquisition protocol, and thus non-linear conditional probabilities seem more appropriate to model them. That is, the probability of an outcome happening when multiple preceding events have occurred is not always equal to the product of the conditional probabilities of this outcome occurring given only one of the preceding events. Learning models therefore must in some manner involve processes of approximating these conditionality relations, such as those seen in non-linear discrimination learning. Exactly how such causal connections are approximated is highly dependent upon how the stimuli themselves are represented by a model. Two benchmarks of non-linear discriminatory performance of a model have been the aforementioned negative patterning and biconditional discriminations. The difficulty, and hence importance, of negative patterning lies in the simple breakdown of linearity on the compound trials. That is, the animal must learn to withhold responding on trials when two cues, which individual predict the outcome, are presented. Biconditional discriminations involve yet more complex non-linearity. Four cues are presented in pairs, as the individual cues offer no information for solving the discrimination.

The RW model, as an elemental model, kept with the assumption of Spence, Konorski, and Estes (Harris, 2006; Wagner, 2008) that sets of elements constituting the attributes of individual stimuli enter into associations with the US directly. As a result, in its original formulation, RW was unable to account for both negative patterning (Rescorla et al., 1985) and biconditional discriminations (Rescorla, 1972; Saavedra, 1975). The model assumes that if each individual stimulus is presented in opposite contingencies, this amounts to partial reinforcement of the cue and hence the discrimination will not be solved. In the case of negative patterning, the model would predict greater responding on the compound trials than

individual trials, due to the summation of the associative strengths of A and B. That is, the model preserves the linearity of summation by positing that the responding elicited by a given configuration of cues is directly proportional to the sum of the individual associative strengths of the constituent stimuli. The addition of elements common to multiple stimuli to RW, which is functionally equivalent to adding a redundant cue X, allowed for the NP discrimination to be solved as the elements common to both A and B acquire superconditioning, while the unique elements of both stimuli become inhibitory toward the outcome. Hence, simply by assuming some similarity between CS representations, it is possible to account for some non-linear discriminations within an elemental framework. In the case of more complicated discriminations such as the biconditional and the tri-conditional discriminations, common elements are not a sufficiently powerful mechanism. The addition of configural elements to the model, which become active during the presentation of a specific compound, allows the RW model to get around this hurdle (Rescorla, 1972). That is, it is assumed that the compound is represented by the animal as more than the sum of its parts. Thus, the NP discrimination is produced in a contrary way as compared to when common elements were used, as configural cues in this case take on inhibitory values toward the US, while the unique elements become excitatory. While configural elements lead to the model predicting deviations from summation linearity in non-linear discrimination training, it nevertheless preserves linearity in the absence of such conditioning. As such, it was unable to account for evidence of linearity of summation being broken (Razran, 1939), that is for configurations of stimuli eliciting more or less response than expected by summing the response strengths elicited by individual presentations of the constituent stimuli. For instance, the observation that the presentation of a stimulus compound, after reinforcing the two stimuli separately, does not always produce more responding (Pearce et al., 2002) contradicted the RW model.

REM model

Such evidence of linearity of response summation being broken, prompted the conception of the replaced elements model (REM) (Wagner, 2003), which manoeuvred around this difficulty by a process whereby some elements of a stimulus are added (configural elements), and some are removed (unique elements), when the cue is presented in a compound. These added and replaced elements are consistent throughout presentations. The added elements correspond to elements representing context-dependent features of the stimulus (i.e. denoting conjunction with another cue), while the replaced elements are assumed to represent features of the cue that are uniquely present when the cue is presented alone. These statistically independent replacement rules are captured by the following relation:

$$P_A(A \cup \text{Comp.}) = \prod_{K \in \text{Comp.}} (1 - r_K) = \prod_{K \in \text{Comp.}} s_K \tag{1.7}$$

Here P_A denotes the proportion of elements of stimulus A that are sampled both when A is presented in isolation and when compound stimulus Comp. is presented. The similarity measure $1 - r_K$ or s_K, denotes the quantity of elements that are sampled on both A and AK trials. Hence, the overall proportion of elements sampled on both A and Comp. trials is the direct product of these binary relations.

The model, as seen in the equation above, postulates a replacement parameter r, which determines what proportion of elements are replaced in this manner. Thus when $r = 0$, the model is equivalent to the RW model, while $r = 1$ produces a purely configural model (Schultheis et al., 2008). The intermediate values allow the model to postulate that summation is not absolute. For instance, $r = 0.5$ leads to the prediction that no summation is observed, that is the compound of two CSs is no more predictive of the outcome than an individual CS. The effect of this replacement is more obvious when one considers the effect

of r upon the quantity of responding elicited by a CS compound AB after A and B have been conditioned to an asymptotic level in isolation. Figure 1.4, $r = 0.0$ leads to perfect summation of the strengths of A and B, while $r = 1.0$ leads to no responding, due to all the elements being replaced. Thus, by assuming that different modalities of stimuli undergo different replacement rates depending on their similarity, consistent with psychophysical theory (Glautier et al., 2010), the model can explain why summation seems to depend on the types of stimuli used (Wagner, 2003).

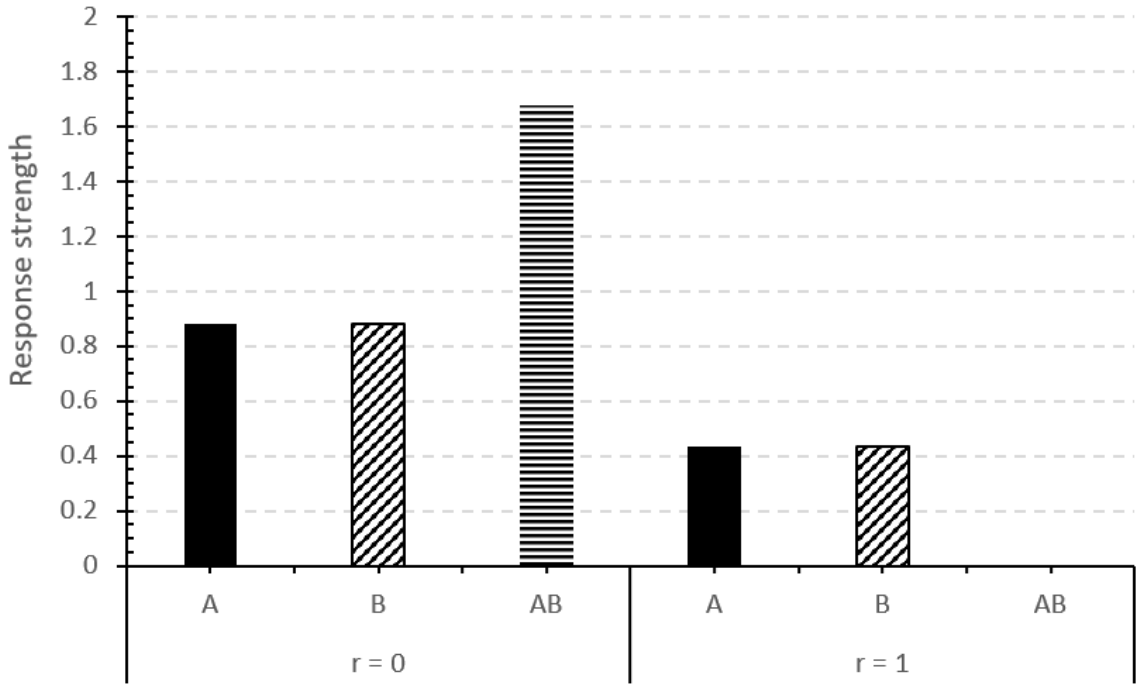


Fig. 1.4 Simulation of the REM model using the REM simulator (Ghorashi, A., Mondragón, E. and Alonso, 2017). Results of the conceptual design 20A+/20B+ (random) followed by test trials for compound AB, with $r = 0$ and $r = 1$.

Though the configural elements representation of RW and the REM model account for some aspects of deviance from perfect summation, they are less capable of dealing with other observed effects. Both models assume that a redundant cue, X, will facilitate the learning of a non-linear discrimination, yet the opposite is often observed (Pearce and Redhead, 1993). A further difficulty is that an underlying implication of most elemental models is

the potential complete reversibility of an association between a stimulus and the outcome should the previously learned contingency be reversed (e.g. reinforcement followed by non-reinforcement). However, experiments displaying retroactive interference in feature negative discriminations, in which B+ trials failed to extinguish the animal's performance on a discrimination A+/AB-, which presumably was learned through B becoming a conditioned inhibitor, contradict this assumption that the response elicited by B is expressed linearly on AB- trials (Wilson and Pearce, 1992). Both the RW model and the REM extension have difficulty reproducing this effect. The former, due to its linear summation, predicts complete interference of B+ training on the A+/AB- training. That is, it expects that after B+ training the AB- compound will display less complete suppression of responding. The latter, though it in principle can reproduce the effect, is forced to postulate a very high replacement rate of B elements. However, retroactive interference in feature negative discriminations has been produced by stimuli from different modalities (Pearce and Wilson, 1991), which seems to require a low replacement rate of elements.

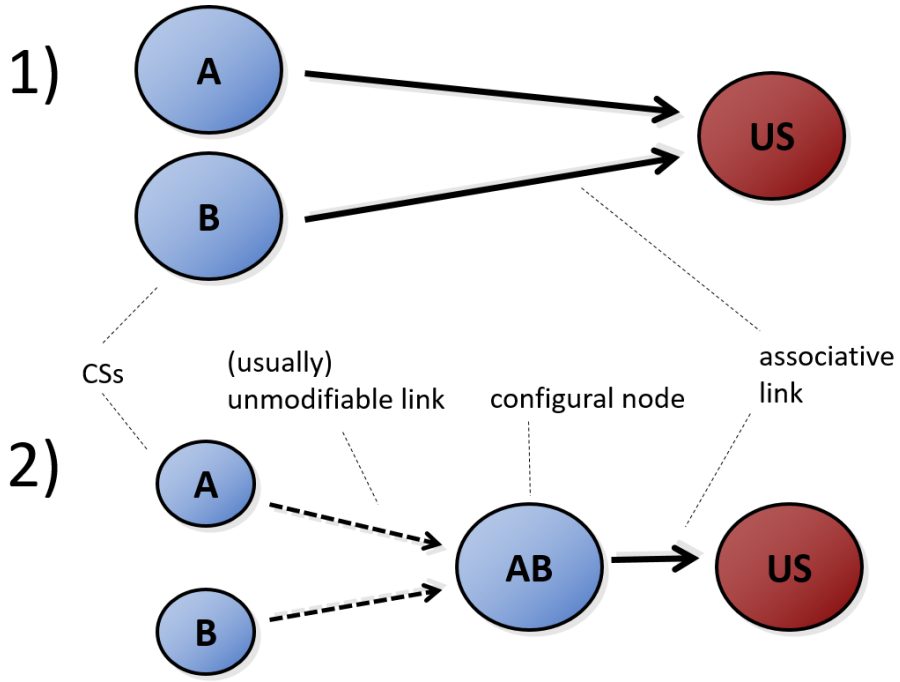


Fig. 1.5 1) An elemental connectionist structure. CSs are directly associated with the occurrence of the US. 2) A configural connectionist structure. CSs activate a configural representation, which in return enters into association with the US.

Pearce model

An alternative account to the elemental approach is supplied by the model of Pearce (Pearce, 1987), which in contrast presumes that nodes representing individual stimuli connect to an additional configural node representing their aggregation (as depicted in segment 2 of Figure 1.5). This node in return forms associations with the outcome.

The learning rule of the model hence involves calculating links between these configural nodes, and utilizes the following link-update equation:

$$\Delta V_i = \beta [\lambda - (V_i + \sum_{i,j} S_{i,j} V_j)] \tag{1.8}$$

Here V_i is the associative strength, β is the US salience, λ is the US asymptote, $S_{i,j}$ is the similarity between stimuli i and j, and the sum ranges over all pairs of CSs.

The similarity measure in return is calculated according to:

$$S_{i,j} = \frac{N_{c_i}}{N_{t_i}} \frac{N_{c_j}}{N_{t_j}} \tag{1.9}$$

Here, N_{c_i} and N_{c_j} represent the number of elements respectively of cues i and j, which are common to i and j, and N_{t_i} and N_{t_j} are the total number of elements of the two cues.

The Pearce model's configural stimulus representation produces many non-linear discriminations simply through the learning between a configural node and the outcome not directly interfering with other configurations. For instance, in Figure 1.6 we have simulated a biconditional discrimination using the Pearce Model Simulator (Gheorghescu, A., Mondragón, E. and Alonso, 2017). Non-linear discriminations are often very difficult for animals to learn, and Pearce's model accounts for this difficulty by postulating that responding to a configuration of stimuli is affected by its similarity to other configurations (as measured by a similarity index). Hence, it anticipates that the difficulty of a non-linear discrimination is directly proportional to the similarity of the constituent compounds of it. In the case of the bi-conditional discrimination, the complexity of the discrimination arises from each compound being similar to another compound undergoing opposite reinforcement (e.g. AB+ and AC-). The model naturally accounts for summation between stimuli not always being observed. Correspondingly, it faces a challenge explaining evidence that summation does sometimes occur. Further, the model predicts a symmetrical deficit in responding (generalization decrement) when a stimulus is added in compound to a previously conditioned stimulus (external inhibition) or when a stimulus is removed from the training compound

(overshadowing). Yet it has been found that overshadowing produces a larger deficit in responding than external inhibition (Brandon et al., 2000). Finally, it has been found that familiarity with the constituent stimuli of a discrimination can facilitate the learning of the discrimination (Hall, 1991; Mondragón and Hall, 2002). This phenomenon of perceptual learning seems to suggest that the animal's perception of the similarity of two stimuli is learned rather than given a priori as through the Pearce model's similarity index. Hence, some form of attentional or learning-based process seems to be necessary to dichotomize unique and redundant elements of stimuli when they are pre-exposed together, such as elements shared by the pre-exposed stimuli losing associability.

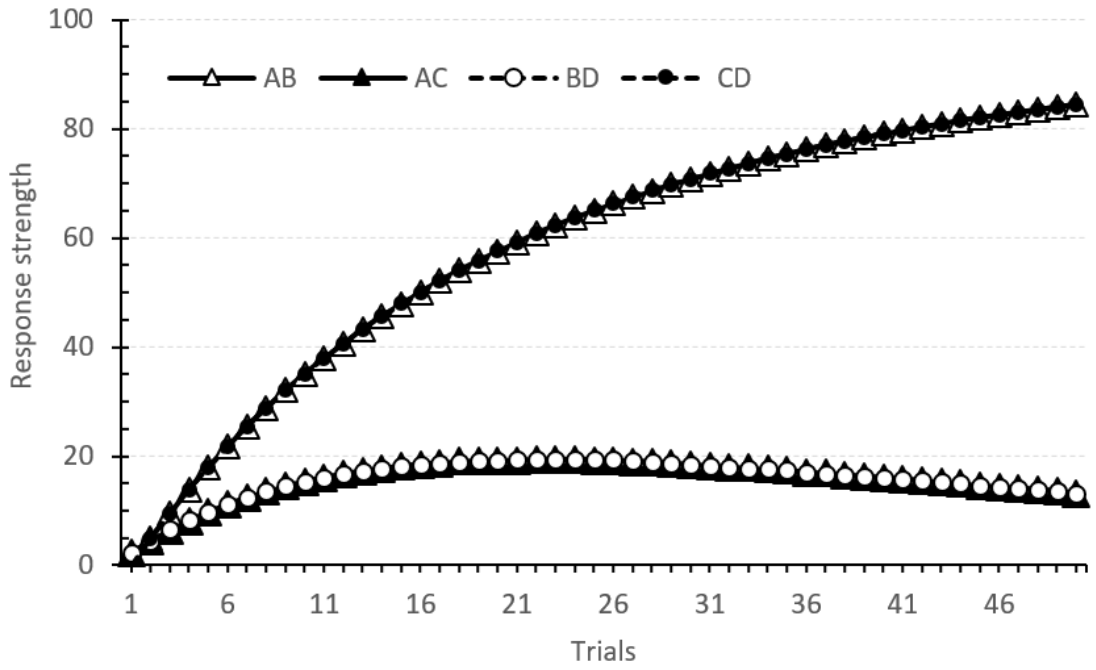


Fig. 1.6 Simulation of the Pearce model using the Pearce Model Simulator (Ghorashi, A., Mondragón, E. and Alonso, 2017). Results of a biconditional discrimination are displayed.

Harris model

Configural models have been further critiqued from a theoretical perspective. In terms of parsimony, they require representations of both individual stimuli as well as configurations, thus making them more complex than purely elemental models (though (Ghirlanda, 2015) has demonstrated that a mapping between elemental and configural representations can be constructed analytically). Furthermore, they seem to take for granted representational information that is usually postulated by purely associative means, namely that a combination of stimuli co-occurred. Finally, they require some limiting process on the quantity of configurations that can form to avoid an infinite generation of different configural representations. On these grounds, and in order to resolve the empirical difficulties outlined, Harris has introduced a purely elemental model utilizing a unique attentional process (Harris, 2006; Harris and Livesey, 2010) to circumnavigate the aforesaid problems of the elemental approach. In the model a finite 'attentional buffer', which makes the activation of elements stronger and more persistent (and therefore contribute more towards responding), is introduced (Figure 1.7). Entry into this buffer is regulated by the change in activation of a given element when it is presented or predicted, and is proportional to the salience of that element (with the saliences of elements of a stimulus assumed to be normally distributed). That is, elements that undergo a strong increase in their activation can enter the attentional buffer. If the buffer is at full capacity, elements with larger changes in activation displace elements with smaller changes in activation. Further, the extent of reinforcement by a US is proportional to the number of US elements that are pushed into the buffer. However, in general, the quantity of CS elements in the buffer does not influence the direction of learning. The buffer simply speeds up learning as the activation of an element functions as a de facto salience.

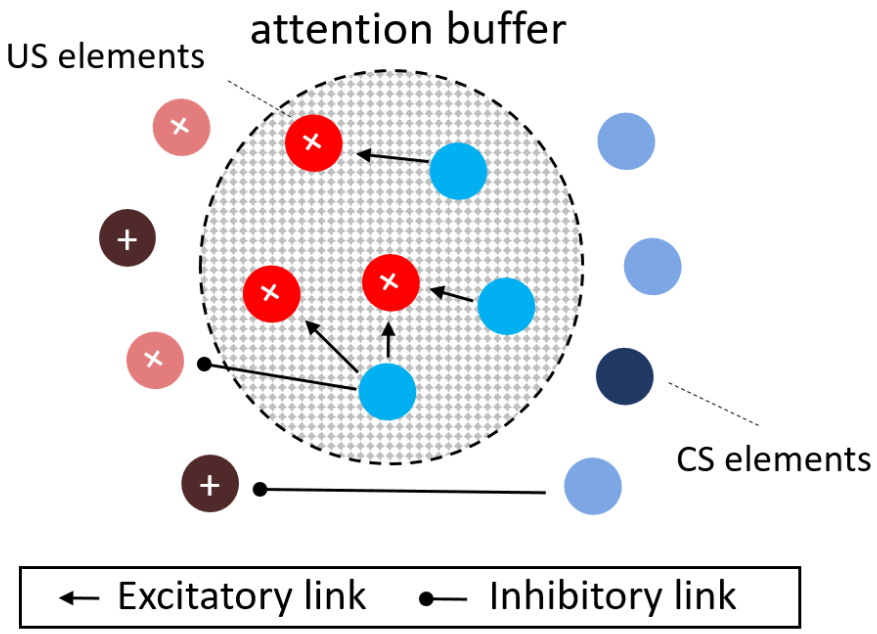


Fig. 1.7 The fixed-capacity attentional buffer of the Harris (2006) model. Elements compete to enter the buffer, with those with a higher weight doing so more readily. The buffer increases the persistence of their activity. The number of US elements inside and outside the buffer determines the overall direction of learning (excitatory or inhibitory).

The entry to the buffer, as mentioned, is regulated by the weight ω_i of stimuli, such that:

$$\sum_i \omega_i \leq \text{ buffer capacity} \tag{1.10}$$

The actual entry of a given element is specified as occurring in proportion to $\Delta\omega_i$, which is the difference between the intrinsic self-weight of a stimulus, and the prediction for it by other stimuli. Since this term acts as a prediction error, it is used in the associative strength update equation of the model:

$$\Delta V_{x,y} = \begin{cases} \omega_x \cdot \beta_y \cdot 2\Delta\omega_y & \text{if } \Delta\omega_x > t \\ \omega_x \cdot \beta_y \cdot (-\Delta\omega_y) & \text{otherwise} \end{cases} \tag{1.11}$$

The response elicited by a stimulus A is thence calculated using the dot product over its elements:

$$R(A) = \sum_i \omega_{A_i} \cdot V_{A_i} \tag{1.12}$$

An important implication of the introduced buffer is that when a stimulus compound is presented, only the most salient elements can enter the buffer. Hence the model can account for partial summation. This partial summation, together with an assumption of common elements equips the model with the capability of predicting various non-linear discriminations. It additionally predicts that these discriminations are solved more slowly when a redundant cue is present, as the redundant cue leaves less capacity in the attentional buffer for relevant cues. The facilitation of discriminatory learning by pre-exposure, i.e. perceptual learning, arises through a process of common elements losing associability through self-prediction hindering their entry into the buffer. Crucially, the model can explain both the lack of retroactive interference in feature negative discriminations (Wilson and Pearce, 1992), as well as the discrepancy between external inhibition and overshadowing. In the case of the former, it postulates that during the B+ presentations, the elements of B that entered the attentional buffer on AB- trials are more inhibitory than the ones which did not. Thus, the elements that did not enter the buffer acquire more excitation than the ones which did. Resultantly, when the AB compound is once again presented, much of the increase in B's associative strength will not be manifested due to the most excitatory elements not entering the buffer. In the latter case, the model explains that in the case of external inhibition, the novel stimulus pushes elements of the compound out of the buffer. However, these elements nevertheless remain active at a lesser level, thus still contributing towards responding. When a cue is removed as in an overshadowing test, its elements are completely inactive, and hence the asymmetry between the phenomena is accounted for. Thus, the Harris model shows that many of the apparent pitfalls of elemental models can be avoided using a richer stimulus representation and a process of selective attention; though the exact nature and substratum of

the attentional buffer remains to be validated by empirical data. Further, many of the unique predictions of the model could be explained also for instance by the simpler assumption of the salience of a cue directly influencing the strength of its activation.

In conclusion, both elemental and configural approaches to stimulus representations have distinct advantages in explaining non-linear discrimination learning. Configural models can offer a solution on how animals learn complicated discriminations, generalization between them, as well as for partial summation. Elemental models must postulate additional learning and representational mechanisms to account for many of these effects. For instance, configural elements or the attentional buffer of Harris significantly extend the reach of elemental analysis. They nevertheless have the advantage of maintaining representational simplicity as well as encapsulating more information in terms of pure associative learning. That is, in many formulations they avoid the problem of how configural representations emerge in the first place. As such, we have favoured this elemental approach for the current model.

Attention and stimulus Pre-exposure: CS processing

As the Harris model assumes that any excitatory link from one stimulus to another induces activation in the recipient stimulus, it can explain various pre-exposure and habituation effects as well. In effect, the associability of a cue, all else being equal, is therefore directly proportional to its novelty. In addition, the model's prediction that the effective associability of stimuli decreases in relation to how many further stimuli are active has been experimentally validated (Lachnit et al., 2008). This sets it in opposition to other paradigms of selective attention.

Mackintosh model

For instance, the Mackintosh model assumes that animals attend to cues that are relatively better predictors of outcomes than other cues (Mackintosh, 1975). It models the attention to a cue, α_i , as changing in proportion to the relative predictiveness of the cue (in relation to other cues) for the outcome. This is modelled through the equations that follow.

The weight update rule is a 'local', non-competitive one, similar to the Hull rule:

$$\Delta V_i = \alpha_i(\lambda - V_i) \tag{1.13}$$

Unlike in the RW model, the cue-competition occurs through changes in the associability α_i. The α_i of a CS rises if $|z\lambda - x_i V_i| < |z\lambda - V_T|$, where V_T is the total associative strength of all other cues, z is the US presence, and λ is the US asymptote. This magnitude of change is postulated to be in proportion to $|z\lambda - x_i V_i| - |z\lambda - V_T|$. The implication is that selective attention is learned, retained for future learning, and presumably aids in reducing proactive interference between stimuli, thereby speeding up learning as discussed in (Kruschke, 2009). Cue-competition effects are thereby explained by purely attentional means and thus the model

only requires the linear operator delta rule familiar from Hull's model. The Mackintosh model's unique assumption that the selective attention paid to a cue has direct reinforcing effects has allowed it to predict phenomena that pose a problem for other models. For instance, it can explain unblocking, wherein the surprising partial omission of reinforcement during a blocking treatment attenuates the blocking of a cue (Dickinson et al., 1976). It thus avoids the prediction of over-prediction pushing the later introduced novel cue towards becoming inhibitory, which the RW model predicts in some circumstances for this treatment. An alternative explanation to that offered by the Mackintosh model for this effect is however that differential reinforcement is represented by the animal as different reinforcers. As such, this result can be accounted for by the RW model. The Mackintosh model also accounts for the learned irrelevance effect (Bonardi and Hall, 1996; Mackintosh, 1973), whereby a CS uncorrelated with US presentations shows poorer subsequent acquisition, due to the best relative predictor accruing the most associability, and hence conditioning more quickly than competing cues. A well-known difficulty faced by the model is however the phenomenon of Hall-Pearce negative transfer (Hall and Pearce, 1979), wherein reinforcing a CS with a weak US, thus making it more a better predictor hinders subsequent conditioning between the same CS and a stronger version of the same US. In this case, it erroneously predicts greater excitatory learning between the previous best predictor of the outcome and the outcome. The model also is unable to account for superconditioning Rescorla (1971b, 2004), wherein presenting a conditioned inhibitor of an outcome together with a novel cue leads to stronger excitatory conditioning to the novel cue than if it were conditioned individually. As the Mackintosh model does not assume a competitive error-term, it cannot account for this effect. The hybrid model of Le Pelley however is able to produce this effect by integrating a combined error-term (Le Pelley, 2004).

Pearce-Hall model

Conversely , the effect of Hall-Pearce negative transfer is easily accounted for by the Pearce and Hall (PH) model (Pearce and Hall, 1980), which postulates that attention rises when the outcome is uncertain, an assumption formalized in the following alpha update rule:

$$\alpha_i^t = |z\lambda^{t-1} - \sum_j x_j V_j^{t-1}| \tag{1.14}$$

where α is once again the associability, z the US presence, λ is the asymptote of learning, and t represents trials.

Excitatory learning in the model is calculated using:

$$\Delta V_i^+ = \alpha_i^t S_i z\lambda \tag{1.15}$$

And inhibitory learning is calculated through:

$$\Delta V_i^- = \alpha_i^t S_i(\sum x_i V_i - z\lambda) \tag{1.16}$$

In both equations S_i is the salience of the US.

That is, the associability of conditioned stimuli rise in accordance to the general uncertainty of the outcome instead of tracking the relative predictiveness of the cue. By implication, in the Hall-Pearce negative transfer effect, the prior training with a weaker US produces a decline in associability which hinders subsequent learning with the stronger US.

In contrast to the Mackintosh model, the Pearce and Hall model has trouble explaining learned irrelevance, as the predictor in a learned irrelevance procedure will accrue more associability due to their lack of correlation with the occurrence of an outcome. In the context of a standard acquisition and extinction protocol the PH model (Pearce and Hall, 1980) predicts a sudden increase in associability when the contingency is changed (i.e. the beginning of acquisition or extinction, Figure 1.8), with a gradual decline thereafter. Further, a great success of the model has been predicting latent inhibition (Channell and Hall, 1983), wherein pre-exposure of a CS attenuates subsequent acquisition training with the same CS. It predicts this effect as arising from the pre-exposure leading to a decline in the associability of the CS, which subsequently rises again during acquisition training. Empirical evidence exists for the hypothesis that prediction errors increase the associability of cues, however a confounding factor is the difficulty of disassociating increases in associability (speed of learning) from direct reinforcement (extent of learning) produced by prediction errors (Holland and Schiffino, 2016). Uncovering and formalizing the precise nature of modulating factors of attention, and their relation to the dopamine system (Ahveninen et al., 2000; Nieoullon, 2002), is hence critical in determining the plausibility of the associabilities of the Mackintosh and PH models. It has been noted that the attentional mechanisms underlying these two models are not necessarily in opposition. The PH attention rule, when instantiated in a real-time model, can for instance produce an increase in associability for the best predictor of the outcome if said predictor produces a prediction error for the US before the US onset. That is, the rule produces more complicated emergent behaviour when instantiated in a real-time model. The existence of evidence supporting each model has fostered the proposal of dual-factor attentional models. One such model is the Le Pelley model (Le Pelley, 2004), which combines both the associability of the Mackintosh model and of the PH model (with the latter given more influence). According to the model, both the general uncertainty of the outcome, as well as the relative predictiveness of a cue determine the overall associability of

the cue. As such, the Le Pelley model can explain phenomena that have proven difficult for either model in isolation. It predicts for instance Hall-Pearce negative transfer as well as the learned irrelevance effect simultaneously. It however introduces further complexity as the two rules can often cancel the influence of one another . Thus, it is also difficult to isolate their respective effects empirically.

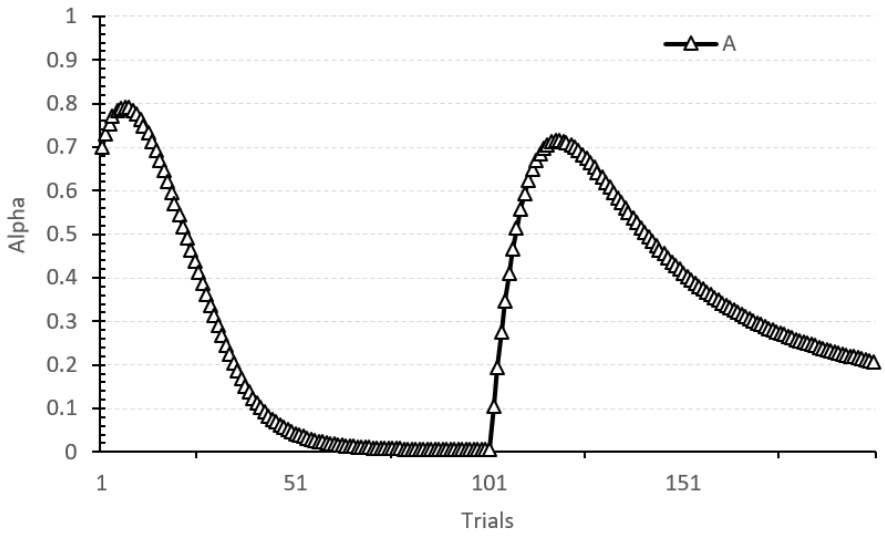


Fig. 1.8 Simulation of the Pearce-Hall model using the Pearce-Hall Model Simulator (Grikietis, R., Mondragón, E., and Alonso, 2016). Alpha value of cue A in a simulation of 100A+/100A- is displayed.

SLGK model, Hall-Rodriguez model

The assumption that novelty of cues is crucial to how learning unfolds, pioneered by the Mackintosh and PH models, was expanded upon by the SLGK model (Kutlu and Schmajuk, 2012). It postulates that the novelty of every stimulus affects its speed of learning towards other stimuli. Further, this novelty is modelled through each stimulus learning to predict each other stimulus through associative links. Therefore, not only reinforced, but also non-reinforced (i.e. 'neutral' or 'silent') learning is incorporated. The model is configural in the sense that it postulates a layer of units intermediate between the sensory units and the

US unit. These configural units have non-modifiable random incoming connections from all CS representations, and their connection to the outcome is modified through a delta rule. Their associability or associative rate is taken to be very low initially, however if the US novelty remains high during training (i.e. the animal is unable to learn a given contingency), then the associability of the configural units rises. It therefore accounts for the difficulty of animals to solve non-linear discriminations. This mechanism of detecting, through a persistent outcome error, when a purely linear approach has failed offers a robust approach for introducing non-linearity into models of conditioning while pre-empting a combinatorial explosion. It can be utilized to change the associability of other representational elements besides configural cues (e.g. common elements) to reflect when the animal cannot solve a learning problem.

Learning between non-reinforcing cues, i.e. 'silent learning' incorporated by the model endows it with the capacity to explain various pre-exposure effects. For example, it explains the attenuated positively accelerating response curve of a latent inhibited stimulus during subsequent acquisition. This response curve further approaches the standard learning asymptote after a sufficient number of reinforced trials (Lubow, 1965). In this regard, it shares similarities with two other models that postulate both silent learning and novelty-based associability, SOP (Wagner, 1981) and the McLaren-Mackintosh (McLaren and Mackintosh, 2000) models. In all three models, latent inhibition emerges from the pre-exposed CS losing novelty by being predicted by the context and itself (unitization). This loss of novelty slows down subsequent acquisition training. A different approach is that taken by the aforementioned PH model and the Hall-Rodriguez extension of it (Hall and Rodriguez, 2010), are also commonly proposed to account for latent inhibition. These predict that the repeated pre-exposure of a cue reduces the attention paid to it, thus reducing its associability. The Hall-Rodriguez model further predicts the formation of a $CS - \neg US$ link, which thereafter

interferes with the response elicited by the CS-US link formed in the subsequent acquisition training (Figure 1.9). This latter mechanism bears a similarity to the explanation offered by the comparator model, which similarly stresses processing during retrieval. Nevertheless, the neutral-learning based account of SLGK, SOP and McLaren-Mackintosh seems to hold a categorical advantage in explaining latent inhibition, as they do not need to assume such a no-US representation or the existence of an extant associative link to it. Hence they avoid predicting inhibitory properties of the pre-exposed CS towards the US, which have not been found in summation tests (Rescorla, 1971a). Further, the assumption of a no-US representation implies a generalization of inhibition to all reinforcers. However, CS-US learning has been shown to be highly US specific, for instance through reinforcer de-valuation effects (Colwill and Motzkin, 1994).

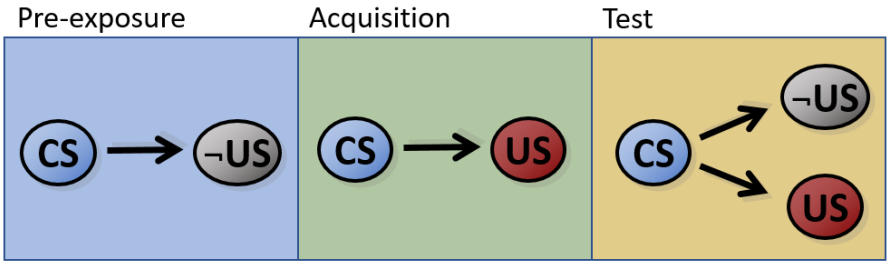


Fig. 1.9 (Hall and Rodriguez, 2010) postulates that pre-exposure induces the formation of a $CS - \neg US$ link, that subsequently interferes with the CS-US link formed during acquisition.

In conclusion, models that assume selective attention is acquired by the most predictive cue, such as the Mackintosh model, offer a robust account of effects such as unblocking and learned irrelevancy. They however are unable to offer a convincing explanation of pre-exposure effects. Similarly, alternative explanations for unblocking and learned irrelevance can often be produced by error-correction effects. In the former case by assuming that differences in US presentations imply differences in the asymptote of learning, that is by assuming that different US presentations increase the effective error term (and hence reduce competition) either by being processed by the animal as separate stimuli to be predicted or

through more intense USs having a higher asymptote within the error term; in the latter case by postulating that uncorrelated CS-US exposures results in inhibitory learning, for instance through the context predicting the US on CS- trials, hence making the CS into a conditioned inhibitor. In contrast, models that postulate a loss in associability as the outcome error declines (PH) or when a cue becomes less novel (SLGK, SOP, and McLaren-Mackintosh) are uniquely capable of explaining both the result and mechanisms behind the loss in associability occurring during pre-exposure. The latter models can also predict the context specificity of LI observed when a pre-exposure is conducted in a context different from that of conditioning. Moreover, when instantiated in a real-time model, the uncertainty-driven variable associability of the PH model could produce emergent Mackintosh-like effects through the best predictor of an outcome producing a strong prediction-error before the onset of the outcome, thus increasing its own associability towards the outcome. We have hence chosen to incorporate a similar mechanism, with important changes involving re-evaluation of persistent uncertainty, to account for learned attentional bias in the Double Error model.

Non-reinforced and Mediated Learning : neutral stimulus associations.

If an animal is learning the contingencies governing events, does it re-evaluate the presumed most probable contingency when it is contradicted by subsequent learning? If so, how? Since it is assumed that prior learning is stored as associations, this revaluation by implication involves the modification of previously formed associations. Behavioural and neural data exists, which demonstrates that a present cue can retrieve the representation of another cue and thereby invoke re-evaluation of the latter's previously formed associations towards a reinforcer (Dickinson, 1996; Holland, 1983; Holland and Forbes, 1982). For instance in terms of neural data, the left hippocampus has been implicated in mediated learning in humans (Jie, 2008) and the dorsolateral prefrontal cortex has been tied to violations of previously formed expectations (Corlett et al., 2007). These so called mediated conditioning effects are of interest as they allow for an understanding of processes governing learning between representations of stimuli which are present and ones which are absent yet associatively invoked. Of cardinal significance to this thesis are the retrospective revaluation effects (Le Pelley and McLaren, 2001) backward blocking (BB) (Urushihara and Miller, 2010), un-overshadowing/retrospective revaluation (RR) (Miller and Witnauer, 2016), as well as forward sensory preconditioning (FSP) (Brogden, 1939), and backward sensory preconditioning (BSP) (Ward-Robinson and Hall, 1996). Explaining these effects with a simple associative learning rule has been elusive as the effects seem to produce opposite learning directions for the retrieved cue in superficially equivalent reinforcement conditions. For instance, in a backward blocking and in a retrospective revaluation procedure, a compound of events AB is first associated to a motivational outcome. Once this association is established, presentations of one of its components alone, e.g., A, follows. While in a BB these single stimulus presentations are also rewarded, in a RR procedure, subsequent A training is non-reinforced. When B is tested afterwards to assess the effect of the most recent training on the previously stabilised association, BB treatment results in B losing associative strength

compared to the first phase of learning, whereas in the RR treatment the excitatory strength of B increases. That is, strengthening the association between one of the elements of the compound and the outcome, results in a reduction in strength of the concomitant stimulus and the outcome, whereas lowering it, raises the associative strength of its associated stimulus.

In a forward and backward sensory preconditioning procedure either a simultaneous AB or serial $B \rightarrow A$ (FSP) or serial $A \rightarrow B$ (BSP) nonreinforced stimulus compound is first presented. In a second phase, A is reinforced. Subsequent test trials show that B, which has never been paired with the outcome, has nevertheless gained associative strength. Thus, an absent cue B undergoes a change in its associative strength as a consequence of the reinforcement treatment given to its associated cue.

One possible explanatory mechanism for these phenomena relies on the idea of mediated conditioning: During conditioning to A, an associative activation of stimulus B linked to A during the initial compound training is produced. This associatively activated representation of B thereupon strengthens or modulates associative connections between itself and other active stimuli. Of note is that the learning between the absent cue and the present outcome in SPC and BB proceeds in opposite directions, an apparent contradiction. Hence, though SPC can be explained as a mediated conditioning effect, BB cannot. Therefore, a robust explanation of mediated learning must take into account the prior learning taking place in SPC and BB.

Mediated learning has been conceptualized as a type of propositional reasoning process (Mitchell et al., 2009). As such, it can be understood through Bayes' rule, which ties the conditional probability of one event E1 given another E2 (that is $P(E_1|E_2)$) as changing in proportion to its converse $P(E_2|E_1)$ times $\frac{P(E_1)}{P(E_2)}$. As an example, in backward blocking

(AB+/A+), the loss of associability by cue B during A+ trials can be interpreted as a direct consequence of this rule. If we assume the associative link should be in proportion to the conditional probability $P(US|B)$, that is the link the animal learns from B to the US is in proportion to the likelihood that B is followed by the US (Equation 1.17), then the reduction of the conditional probability $P(B|US)$ during A+ trials will lead to a concomitant reduction in $P(US|B)$:

$$P(US|B) = \frac{P(US)P(B|US)}{P(B)} \tag{1.17}$$

A similar argument extends to other mediated learning effects. Mediated learning can further be (in some cases), envisioned as a form of logical reasoning. For instance, in the case of backward blocking again, one can view the first phase of excitatory learning as instantiating an expectation for the logical proposition $B \rightarrow US$. Here $\rightarrow$ stands for logical implication. Thus, if we take the re-evaluation phase of such a design to imply the propositions $\neg B$ and US, that is B does not occur yet the outcome does, then the logical implication is $\neg(B \rightarrow US)$, as otherwise a contradiction would occur. Hence a principled account for mediated learning should be assumed to offer learning or retrieval processes which are guided by such considerations of conditionality of events, and could possibly underlie higher-order reasoning as displayed by animals and humans.

Though the SLGK model forsakes parsimony through some partly redundant psychological processes and by its own evaluation accounts for 94% of learning phenomena, it (by admission of the authors) fails to reproduce some mediated learning effects, specifically mediated conditioning (Kutlu and Schmajuk, 2012). The reason for this failure is not elaborated upon other than a mention that adding λ to the US prediction equation of the model could possibly reproduce the effect.

McLaren-Mackintosh and SOP models

As mediated conditioning is one of the phenomena we are most interested in for reasons that will be detailed later, we will elaborate on how the McLaren-Mackintosh model (McLaren and Mackintosh, 2000) and SOP (Wagner, 1981) extensions account for it and other mediated learning effects.

The McLaren-Mackintosh model is a real-time elemental learning model incorporating error-correction and neutral cue learning. Stimuli are taken to be represented by sets of mutually overlapping elements with time-dependent activation. Uniquely, the model postulates that associative learning is integrated in associative links both on a short and long-term basis. In the short-term, an associative link is thought to increment towards the direction of learning. Subsequently, through a mechanism of weight-decay, the associative link approaches an intermediate value between its original value and the short-term increment. This mechanism furnishes accounts of spontaneous recovery and the Espinet effect (Espinet et al., 1995). Each node in the connectionist network introduced in the model has connections to all other units, and thus has both an external (sensory) input, as well as a modifiable internal (associative) input, both of which contribute to its level of activity. The difference between these two inputs is considered the prediction error for that node, and thus determines the amount of associative learning from other nodes to it (i.e. functions like a standard outcome error-term), as well acting as a modulator of the associability from that node to other nodes, by modulating the level of activity of the node. The complete equations of the model are as follows:

$$
\begin{aligned}
d\Omega_i/dt &= \begin{cases} E(e_i+i_i)(1-\Omega_i)-D\Omega_i & for\ \Omega_i \geq 0 \\ E(e_i+i_i)(1+\Omega_i)-D\Omega_i & otherwise \end{cases} \\
dw_{ij}/dt &= S\Delta_i\Omega_j - Km_{ij} \\
dm_{ij}/dt &= S\Delta_i\Omega_j - Lm_{ij} \\
i_i &= \sum_j w_{ij}\Omega_j \\
\Delta_i &= e_i - i_i \\
modulation &= \begin{cases} r\Delta_i & if\ \Delta_i > 0 \\ 0 & otherwise \end{cases}
\end{aligned}
\tag{1.18}
$$

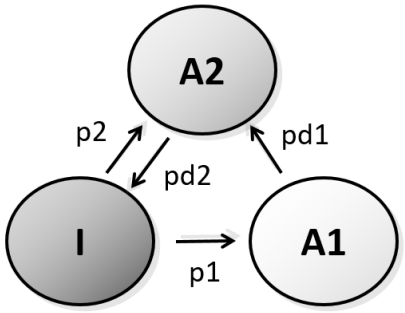
Here Ω_i denotes the overall activity of a unit, w_{ij} is the weight from j to i, S,K,L,r are constants, e_i is the external input of a node, i_i is the internal input to a node, and the modulation is a multiplier of the activity.

The internal input is precisely what allows for the retrieval of absent cues and therefore mediated learning in the model. The model postulates that in sensory preconditioning, the pre-exposure of the compound AB results in the formation of bi-directional excitatory associations between the two CSs. Subsequently, when A is presented together with reinforcement, A associatively retrieves the representation of B. This associatively retrieved representation in return is capable of supporting learning towards the present outcome. Therefore, mediated conditioning is treated equivalently to learning between present stimuli. An equivalent argument holds for backward sensory preconditioning. The model however faces self-ascribed uncertainties in whether it can account for the re-evaluation effects of backward blocking and retrospective revaluation. For instance, in the case of backward blocking the delta rule

employed would suggest that the retrieved cue B could potentially gain additional excitatory strength as its prediction for the outcome would be lower than if it were directly present. Hence the delta term of the outcome would support further learning than in the previous phase. Therefore, to obtain backward blocking, it seems that the model would need to assume, for instance, that present and retrieved cues contribute towards predicting the outcome in proportion to the link strength between the predicting cue and the outcome, that is without multiplying the link strength by the cue's activation.

An early success amongst real-time elemental models was SOP (Wagner, 1981). It aimed to explain both qualitative differences between responses to a stimulus before and after conditioning (i.e. differences between the UR and CR), as well as preconditioning effects wherein a loss in responsiveness to a pre-exposed stimulus, either a CS or US, is observed. For instance, conditioned diminution of the UR is often found when a CS is presented preceding a US (Donegan and Wagner, 1987). SOP predicts this effect through the CS priming the elements of the US, which in return evokes less unconditioned responding than without being primed. The model works through compartmental memory systems in the form of a short-term memory (STM) and a long-term memory (LTM) store. Inactive (I) stimulus representations are activated from long-term storage into short-term memory. These representations are assumed to re-enter the LTM store once they leave the STM store. The STM store consists of an elemental network of processing units consisting of elements in primary (A1) and secondary (A2) states of activation for each stimulus centre/unit, as seen in the top left corner of Figure 1.10 below. The difference between the primary and secondary state is in the degree of activity. The former leads to a state of 'rehearsal' or especially active representation, while A2 activation is weaker and operates at the periphery of attention. As with previous models such as RW, it is assumed that the total quantity of all stimulus A1 elements is limited. Therefore increasing the number of A1 elements increases the speed at

which these elements decay into the A2 state. The decay of A1, $pd1$, is thus assumed to be higher than that of A2 elements, $pd2$. Inactive elements of a unit are activated into the A1 state through stimulus presentation, however this does not occur if the element is already in an A2 state of activation. The A2 state of activation can be produced through either decay from the A1 state, or through the spread of activation among associative links. Lastly, over time elements in the A2 state of activation decay back into the inactive state. The model specifies that the direction of learning (excitatory or inhibitory) depends on the ratio of A1 to A2 elements of the outcome. Thus a more novel outcome supports more learning, whereas an over-predicted outcome supports only inhibitory learning. These processes are summarized as learning rules in the top right corner of Figure 1.10.



			US	
			A1	A2
CS	A1	FW	CS → US	CS ⊣ US
		BW	US → CS	US … CS
	A2	FW	CS … US	CS … US
		BW	US ⊣ CS	US … CS

			US	
			A1	A2
CS	A1	FW	CS → US	CS ⊣ US
		BW	US → CS	US → CS
	A2	FW	CS → US	CS ⊣ US
		BW	US … CS	US … CS

			US	
			A1	A2
CS	A1	FW	CS → US	CS ⊣ US
		BW	US → CS	US ⊣ CS
	A2	FW	CS ⊣ US	CS → US
		BW	US ⊣ CS	US → CS

Fig. 1.10 Top left: SOP activation states: I, A1, and A2 are respectively inactive, active, and associatively activated or decayed activation states, and p1 and p2 are respectively rates by which inactive elements can be activated into A1 or A2 states. pd1 and pd2 are respectively rates at which A1 elements decay into the A2 states, and A2 elements decay into the inactive state. Top right: SOP learning rules. Bottom Right: Dickinson & Burke SOP learning rules. Bottom Left: Holland SOP learning rules.

In SOP, the generation of a response R_j is produced through a weighted function of the A1 and A2 activation of an outcome:

$$R_j = f(W_{1,j}p_{A1} + W_{2,j}p_{A2}) \tag{1.19}$$

where p_{A1} and p_{A2} denote the proportion of outcome elements being in the A1 and A2 state of activation, respectively. The weighting of the outcome's A2 activation in this function may sometimes be negative, resulting in opposing response dynamics.

SOP Learning Rules

The learning rules of SOP specify that the temporarily varying activation of these units results in excitatory or inhibitory links forming between CS representations and the US representation. Excitatory learning between a CS and US is given by:

$$\Delta V^{+}_{CS-US} = L^{+}(p_{A1,CS} \cdot p_{A1,US}) \tag{1.20}$$

It is proportional (according to a learning rate $0 \leq L^{+} \leq 1$) to the number of CS A1 elements active at a given instance of time, $p_{A1,CS}$, multiplied by the number of US A1 elements active at the same time, $p_{A1,US}$. This corresponds to the top left quadrant of the SOP learning rules displayed in the top-right corner of Figure 1.10, *i.e.* excitatory learning occurs when memory representations co-occur in the active A1 state.

Conversely, inhibitory learning between a CS and US is given by the equation:

$$\Delta V^{-}_{CS-US} = L^{-}(p_{A1,CS} \cdot p_{A2,US}) \tag{1.21}$$

It is proportional to the product of active CS A1 and US A2 elements. This will tend to occur when the US occurs and is followed by A2 activation of the CS, or when the CS occurs, followed by A2 activation of the US as seen in top-right corner of Figure 1.10 The net sum of learning is the temporally summed difference between these two quantities, $\sum(\Delta V^{+}_{CS-US} - \Delta V^{-}_{CS-US})$. That is, conditioning is excitatory when the CS and US representations are concurrently at the centre of attention, and inhibitory when the CS is at the centre of attention yet the US is at the periphery.

Predictions of SOP

The SOP model produces acquisition through the excitatory effects of the concurrent activation of CS and US A1 elements. This acquisition reaches an asymptote due to elicitation of US A2 processing by the CS becoming progressively higher as the associative connection from the CS to the US becomes stronger. The process reaches a point where the number of A1 and A2 elements of the US is equal, resulting in portions of inhibitory and excitatory learning which cancel out. If the CS then is put through an extinction procedure, the US A2 processing evoked by the CS results in inhibitory conditioning without any concomitant excitatory conditioning, resulting in a loss of associative strength.

Cue competition originates equivalently from the influence of A2-processing. If a novel cue were to be introduced and presented in conjunction with the CS which has reached asymptotic conditioning, it would undergo blocking. The elements of the US would be elicited into co-equal portions of A1 and A2 activation, resulting in a net conditioning of zero for the novel cue. A similar outcome results when two CSs of unequal salience are

reinforced together. The more salient CS is more associable, and therefore forms a larger excitatory connection with the US. This results in a higher portion of US A2-processing occurring sooner than if the less salient CS would have been conditioned individually. Thus the less salient CS reaches a lower excitatory associative strength before the proportion of US A1 and A2 activation is at a level where no further conditioning occurs.

The most salient feature of SOP is its ability to predict short-term habituation effects (Whitlow and Wagner, 1984), wherein the presentation of a stimulus attenuates its associability in subsequent conditioning procedures due to the inability of stimulus elements which have decayed to the periphery of attention to become reactivated into a more active state. For instance, both CS and US pre-exposure prior to acquisition trials result in a reduction in the observed conditioning. Habituation, according to the model, is the result of stimulus pre-exposure increasing the number of elements active in the A2 state. When the acquisition trials then occur, some of these A2 elements will still be active, resulting in a lower proportion of inactive elements which can be activated into the A1 state. This results in both inhibitory learning due to the active A2 elements, and lack of excitatory learning due to a lesser quantity of A1 elements. Pre-exposure induced habituation is therefore used as the underlying explanation for why CS pre-exposure results in latent inhibition.

SOP and its extensions (CSOP (Wagner and Brandon, 2001), AESOP (Brandon et al., 2003)) can explain conditioning between present and absent cues (Figure 1.10, top-right quadrant), but did not offer rules for doing so between two absent yet associatively activated stimuli. Learning rules for non-present (A2) CSs were introduced by models based on SOP, such as those of Holland (Holland, 1983) or Dickinson and Burke (Dickinson, 1996). The former (Figure 1.10, bottom right quadrant), accounts for mediated conditioning and predicts that two stimuli in the A2 state undergo conditioned inhibition. The latter (Figure 1.10,

bottom left quadrant) predicts backwards blocking as well as mediated conditioning between two A2 stimuli.

As is evident, these rules are in multiple instances (e.g. A2 CS A2 US and A2 CS A1 US rules) the reverse of one another. Surprisingly, experimental data has been collected to support each (Dwyer, 1999), leaving the problem unresolved from the point of view of these two SOP extensions. A further problem is that the proposed learning rules do not seem conductive to being tied to any reasoning process, such as probabilistic propositional reasoning based on Bayes' rule. That is, the psychological meaning behind why A2-A2 learning should be either excitatory or inhibitory is unclear. If both retrospective revaluation and backward blocking are to be explained purely through learning rules, what seems to be needed is therefore a learning rule which subsumes both sets of rules. This supra-rule should reverse the direction of learning between present and absent cues in a principled manner depending on the prior reinforcement history. Further, as the distinction between absent and present cues is not always black and white, it is not apparent what form of mediated learning should occur in situations where the expectation for a cue is of such extent that the animal treats it as being for all intents and purposes present. Therefore, a model of classical conditioning able to account for a more comprehensive set of mediated conditioning phenomena is relevant to advancing understanding of the processes underlying mediated conditioning, and we aim to offer a formulation of precisely such a model.

Other models of mediated learning

It is however possible to explain mediated learning through mechanisms different from alterations in the associative learning rules. The comparator model for instance produces retrospective revaluation through the competition of associative links during retrieval. That is the $B \rightarrow US$ link is relatively larger than the $A \rightarrow US$ link after the latter has lost associative strength in the second phase. However an analytical analysis of various comparator models has cautioned that such re-evaluation effects might be harder to reproduce than originally conjectured (Ghirlanda and Ibadullayev, 2015, p. 18). The APECS model (McLaren, 1993), is a general model of learning (with a strength especially in sequential learning), and is capable of accounting for retrospective revaluation through a hidden layer that recruits units for each novel input-output mapping. Though the TD model does not presume learning between neutral stimuli, the Predictive Representations (Ludvig and Koop, 2008) extension of it accounts for mediated conditioning and backward sensory preconditioning by proposing that present cues retrieve representations of stimuli they have previously presented with. These retrieved cues condition equivalently to present cues, save for having lower levels of activation. The formulation of the model however seems to preclude explaining backward blocking without further assumptions, as learning between a retrieved CS and a present outcome will tend to be excitatory. Hence the model conceptualizes mediated learning as a form of reasoning about causal relations between events, yet does not seem to capture reasoning occurring when prior learning is contradicted by subsequent learning in such a manner that previously formed links should be weakened.

A unique account of mediated learning is conceived of in the 'replayed experience' model of Ludvig, Mirian, Kehoe, and Sutton (Ludvig et al., 2017). It in contrast postulates that the animal re-processes previous learning during the inter-trial interval (ITI), thereby consolidating contradictory learning contingencies (Figure 1.11). It can account for a wide-variety of

mediated learning effects including mediated conditioning and backward blocking in this manner. In a sense, it accounts for mediated learning by postulating that specific contingencies are intermixed as replay experiences in the mind of the animal, and for instance backward blocking is therefore explained in the same way as intermixed standard blocking. However, the model relies on the assumption that such post-hoc revaluation indeed occurs during the ITI, and must assume that representations of different outcome-contingencies have already been learned by the animal for them to be replayed.

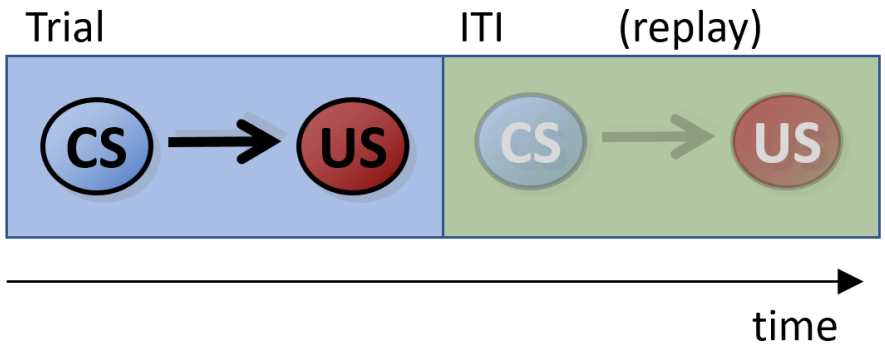


Fig. 1.11 Replay model of (Ludvig et al., 2017). Previously experienced stimulus contingencies are replayed by the animal during periods of reduced stimulation, such as during an ITI.

In summary, a plethora of models have been developed to account for wide varieties of phenomena. Nevertheless, the sets of phenomena explained by models are quite distinct. The pre-exposure effect of latent inhibition is accounted for by a variety of models, however the accounts given by the SLGK, McLaren-Mackintosh, and SOP models excel due to their principled explanation of latent inhibition stemming from a loss in the speed of association formation due to CS elements being predicted both by themselves and the context. In terms of mediated learning, most of the models incorporating neutral stimulus associations can account for some but not all effects, with different models displaying different strengths. The experience replay model can explain all the discussed mediated learning effects through its unique replay mechanism. Yet, in terms of models which rely on learning rules to explain mediated learning, backward blocking and mediated conditioning are not accounted for

simultaneously as they require $A2 \rightarrow A1$ learning to proceed in opposite directions based on prior learning. Similarly, solving various non-linear discriminations, such as negative patterning and the perceptual learning effect, seems to necessitate common or configural representations along with some process whereby their unique features become more associable through pre-exposure. We will now proceed to introducing a new model, which aims to account for all the highlighted classes of learning effects through a general framework of real-time error-correction learning integrating a unique second error-term denoting the novelty of an outcome predictor as well as a dynamic asymptote of learning.

1.6 Methodology

To simulate the results of the Double Error model, we have created an open-source Java simulator for inputting, processing, simulating and producing interactive graphical and data outputs for psychological experiments (Figure 1.12). This simulator incorporates a number of convenient features. These include output graphs for relevant measures of learning as well as model internals, the ability to save simulation designs, and randomization of trial ordering. Designs are entered in the standard manner they are written in learning theory (e.g. $5A+/5AB-$), and various properties of simulated stimuli and other aspects of a design can be modified (duration, onset, intensity, salience, ITI, etc.)

The model code is available on GitHub at: https://github.com/CAL-R/DE_model and https://github.com/nomadicmonad/DE_model

The executable simulator is available at: https://cal-r.org/index.php?id=DE-sim

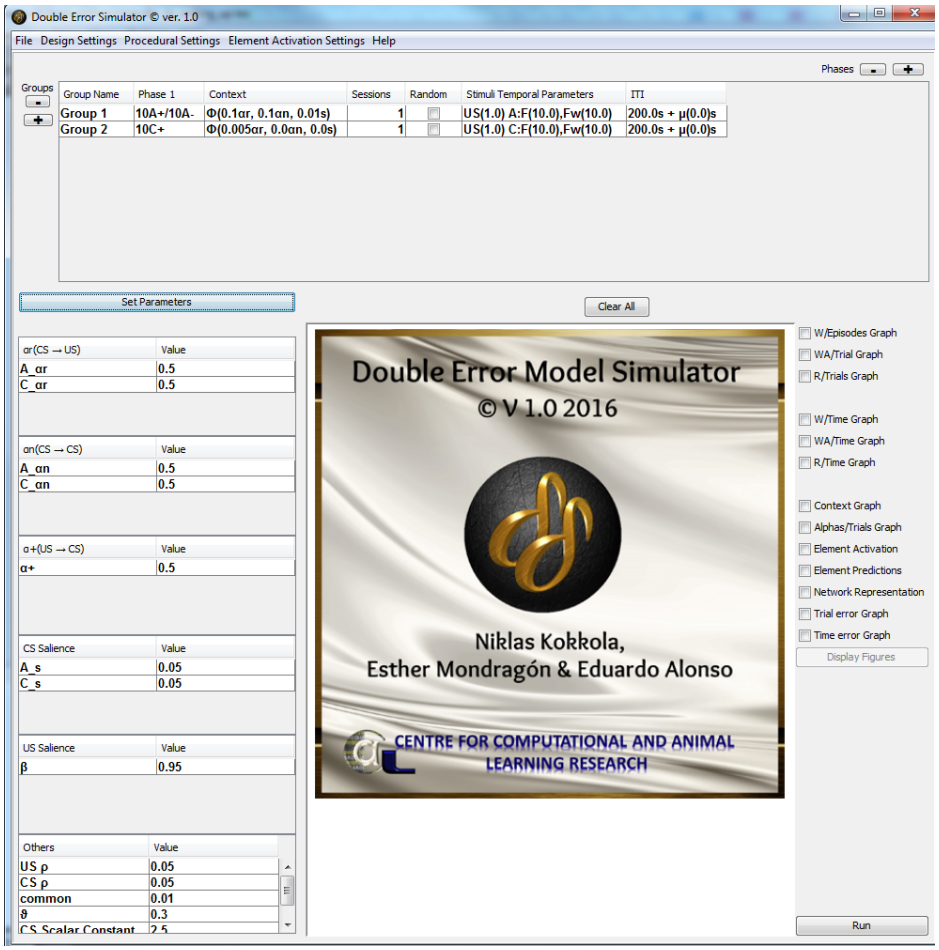


Fig. 1.12 The Double Error Model simulator GUI.

Simulation outputs are stored as excel files, and these files contain all relevant information from a given simulation. This includes the list of parameter values used, changes in all free parameters of the model, time and trial dependent changes in weights, activities and other model values, and values derived through further processing (e.g. responses).

The performance and behaviour of the Double Error is compared to empirical results and extant models both analytically and quantitatively. The function of the model has been reviewed in relation to existing learning theory. Some general mathematical convergence properties of the model have been explored, and its relation to machine learning theory has

been surveyed.

Quantitatively, a wide range of learning phenomena are simulated in the results section. Where applicable, correlation coefficients (R) are calculated for the fit of the simulation to the outputs of a learning experiment. The formula used is simply $R(A,B) = \frac{\mathrm{COV}(A,B)}{\sigma_A \sigma_B}$, where COV is the covariance, and σ is the respective standard deviation. When required for the simulation of an empirical experiment, suppression ratios (r) are calculated using the convention $r = \frac{100-CR}{(100-CR)+100}$, where CR is the trial-based conditioned response generated by the model through aggregating real-time values. In cases where response strengths/predicted responding is used as a measure, this corresponds to the trial-based CR, of which the equation is covered in the model section of this thesis.

In the classifier comparison of the Double Error model to a naive Bayes classifier and a perceptron, the error measure used was the mean squared error (MSE):

$$\mathrm{MSE} = \frac{1}{N}\sum_{i}^{N}(\mathrm{target} - \mathrm{prediction})^2 \tag{1.22}$$

In the natural language task, the measure used is the ratio between the expectation the model created for the next alphabetic character to appear $(1 - \mathit{prediction})$, and the total level of prediction for all alphabetic characters on that time-point. This ensures the model does not perform well on prediction by over-predicting all possible outcomes.

Chapter 2

The Double Error Model

2.1 The Model

The Double Error (DE) model introduced in this chapter conceptualizes a formal, computational model of classical conditioning. Its ontology is based around a connectionist network of elements with adaptive weights. It utilises an error-correction learning framework, thereby reproducing cue-competition effects predicted by extant error-correction models. Likewise, activation states of elements (directly active vs. retrieved) and accompanying decay dynamics are used, hence predicting time-dependent effects such as differential responding early and late during a CS presentation. Learning occurs at the level of elements, which can be unique to a stimulus or shared between two stimuli. The model incorporates learning between neutral (i.e. non-US) stimuli (including elements of the same stimulus), thus explaining the influence of learning between non-US stimuli on learning between a stimulus and the US. Thus, it accounts for pre-exposure effects such as latent inhibition. Its associative learning rule includes a prediction error-term for both the predicting and predicted stimuli, with both influencing the associability of a CS with an outcome. That is, a more novel predictor will undergo faster conditioning toward an outcome. This so called double error learning rule, along with the unique asymptote of learning used (that measures predictor-outcome co-activation),

endows the model with the capability of accounting for apparently contradictory mediated learning phenomena. It does so by positing that the crucial factors influencing the direction of mediated learning are the strength of retrieval of the absent cue and the prior strength of the link between the retrieved cue and the outcome. Finally, a process governing changes in attention-based associability of reinforcers and non-reinforcers is introduced, and operates based on the time-discounted uncertainty in the occurrence of these classes of stimuli. These attentional processes contribute towards the model's capability of predicting a wide range of phenomena including pre-exposure, non-linear discriminations and revaluation of correlation in learned irrelevance. These predictions of the model are compared directly to empirical results in Chapter 3 of this thesis, showing that the model can reproduce a wide variety of simple learning effects, cue-competition effects, attentional-effects, pre-exposure effects, mediated-learning effects, non-linear discriminations, generalization effects, and timing related effects.

List of Mathematical Notation in the Model

$\mathscr{A}$	total activation of an element
$\hat{\mathscr{A}}$	normalized total activation of the element
α_n	revaluation alpha: speed of learning toward non-reinforcing stimuli
α_r	revaluation alpha: speed of learning toward reinforcing stimuli
b	backwards discount: multiplies asymptote in backward learning
c	wave-constant: standard deviation of a time-wave curve
CV	Linear time-wave width multiplier.
$\delta_{j,p}$	predictor error-term (double error namesake)
$\delta_{j,p\rightarrow m,o}$	outcome error-term from the point-of-view of the predictor
$\delta_{\text{US/CS}}$	moving-average of prediction errors for USs or CSs
$\hat{\delta}$	alpha decay factor
e	eligibility modulator of learning
j,p	jth element of a predictor stimulus
$\lambda_{j,p\rightarrow m,o}$	asymptote of learning to the outcome from the point-of-view of the predictor
m,o	mth element of an outcome stimulus
o	outcome (stimulus/element)
p	predictor (stimulus/element)
s	skew parameter
t	current time-point from the start of the trial
ϑ	associative retrieval/prediction discount
w	associative weight of an element
Y	direct (sensory) activation of an element
$\hat{Y}$	associative activation of an element

Stimulus Representation

The DE model, following the argument of parsimony in the literature review, uses an elemental framework of learning. In the model, a stimulus is constituted by a set of elements that are related to the physical attributes of the stimulus. These elements can be unique to the stimulus, or shared with other stimuli. These common elements (Figure 2.1) are probabilistically sampled whenever one of their 'parent' stimuli is active. These common elements are not shared in common with further stimuli, though this could be done. However introducing such complex dependencies relies on making numerous assumptions about the representation of a stimulus, which we wish to avoid. A more principled approach to building representations of stimuli with commonalities could rely on deep connectionist structures, such as deep neural networks (Mondragón et al., 2017), to learn more neurobiologically plausible representations than ad hoc constructs.

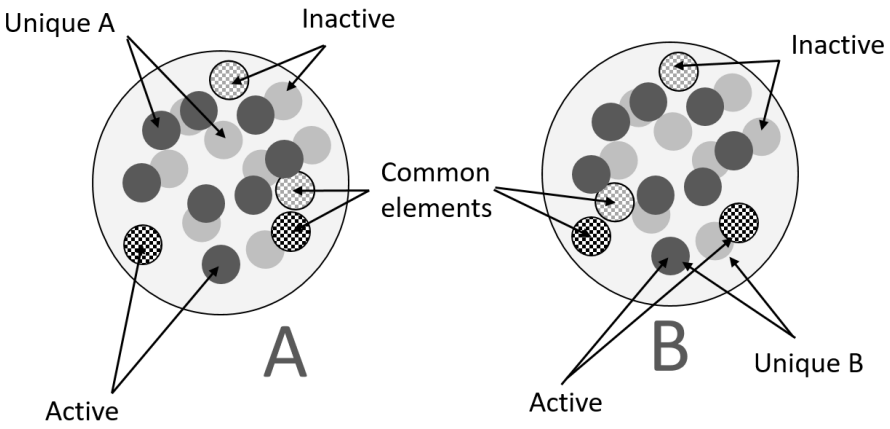


Fig. 2.1 A stimulus is constituted by elements unique to it, common elements shared with other stimuli, and the elements respective activation state.

Each element of each stimulus is capable of developing associative links to each other element (including elements of the same stimulus). Therefore, the model's ontology presumes an elemental connectionist network structure (Figure 2.2 below) of the total stimulus repre-

sentation, with bi-directional (not necessarily symmetric) links between each pair of elements.

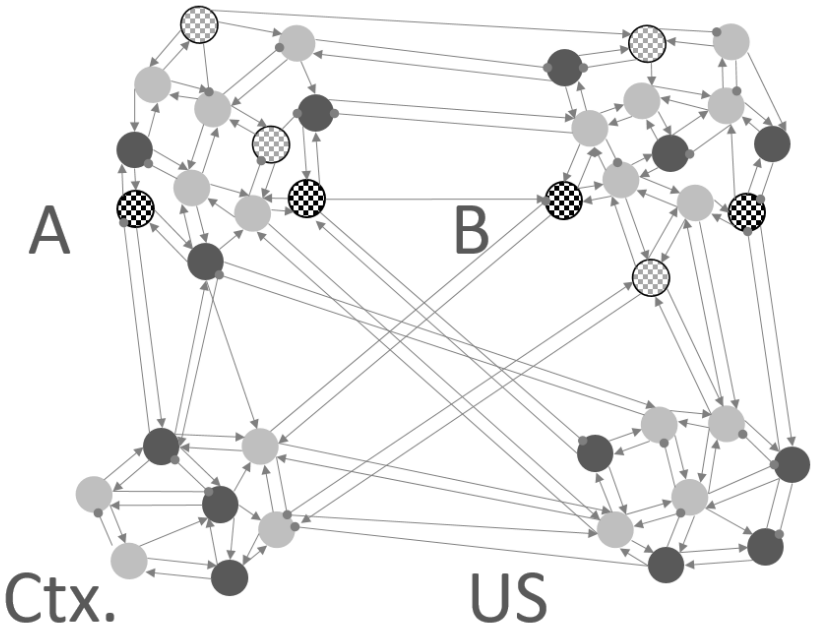


Fig. 2.2 Conceptual representation of the elemental network structure of the DE model.

It is assumed that the stimulus representation varies through time, with various elements being differentially active early or later on during the presentation of a stimulus. Hence, elements form clusters that are differentially active throughout the stimulus presentation, which we term 'time-waves'. A stimulus is set to have one time-wave for each time-unit of duration of the stimulus. Each time-wave of a given stimulus follows an 'approximately' Gaussian distribution, with $Y^t_{j,p}$ denoting the probability that a given element belonging to the time-wave will be directly activated at that moment in time:

$$Y^t_{j,p} = exp(-\frac{(t-j)^2 s}{2c^2 \cdot \mathrm{CV}}) \tag{2.1}$$

where CV is a linear multiplier, $c = \sqrt{\frac{\mu}{\sigma^2}}$ where $\frac{\mu}{\sigma^2}$ is henceforth denoted 'wave-constant', and s is a skew parameter which multiplies the enclosed term when $t < j$. Here j is the mean of the corresponding time-wave.

The curve is not normalized by the usual $\sqrt{2\pi c}$ factor: we desire the peak of the curve to have a maximal value of 1, as this works with the asymptote used in the error-term of the learning equation to allow elements sampled from different time-curves to achieve the same level of conditioning. A time-wave's shape is approximately of the form presented in Figure 2.3 below.

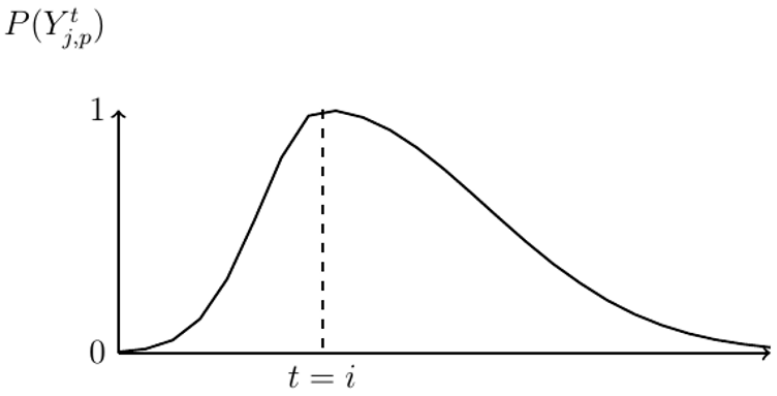


Fig. 2.3 Time-waves are skewed to the right, have a specified peak t and denote the probability that a given element belonging to this temporal cluster is active at a given time-point.

For instance, with a 10 time-unit stimulus, the constituent time-waves of the stimulus correspond to the values in Figure 2.4.

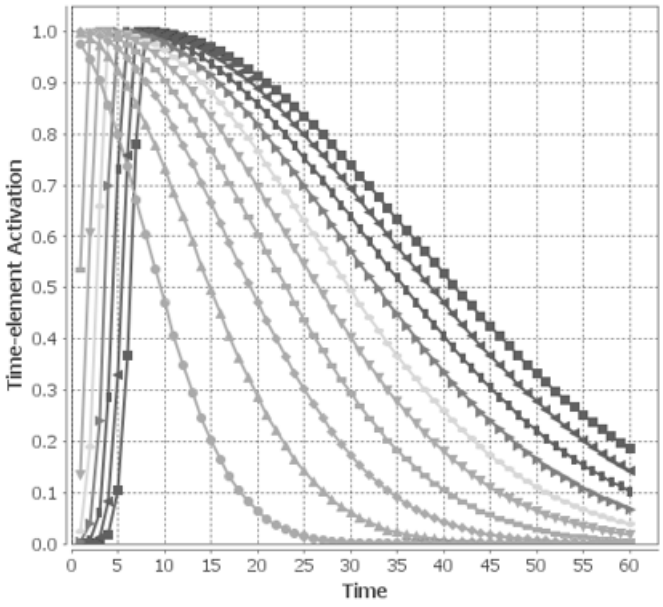


Fig. 2.4 The presence of a stimulus produces a cascade of time-waves in proportion to its length. Each time-wave peaks at a subsequent time-point and its standard deviation is proportional to its mean. A given time-wave gives the activation probability of elements 'belonging' to the wave.

Time-waves are used to represent the idea that the stimulus representation varies in time, so that its elements are differentially active during the presentation of a stimulus. This

differential activity encapsulates principles of both sensitivity to time (time-dependent receptive fields), and noise in predicting time-dependent events (most activity occurring around the mean, with activity in both directions falling off rapidly). The differential activity of time-waves means earlier time-waves elements can learn differential associations from later ones. Similarly, significant generalization through time occurs as the tails of the time-wave activations overlap, allowing the model more flexibility than afforded by step-functions or non-generalizing stimulus activity functions as seen in the original TD model or its CSC extension. These time-waves are hence closely related to the 'micro-stimuli' representation of TD (Ludvig et al., 2009), which was the starting point for my work on temporal representations of stimuli. The persistence of the time-wave activations further allows for effective trace conditioning even after the offset of a stimulus. Thus, the DE model does not rely on separate eligibility traces to be able to produce trace conditioning as in some other models of learning, though it uses a type of eligibility to produce more optimal anticipatory CR curves (which is not strictly needed).

Associative Activation

Whenever a given element is active (including associatively active), it produces predictions for other elements in proportion to the weights from itself to these elements. This amounts to a process of associative retrieval if the predicted element is not directly active. This retrieved element can in return retrieve further stimuli, producing a chain of associative retrieval (Figure 2.5).

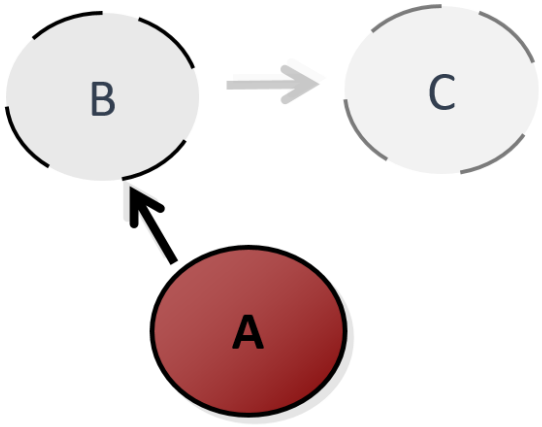


Fig. 2.5 Associative links between stimuli allow stimuli to retrieve the representations of absent stimuli. This can produce chains of associative retrieval, e.g. $A \rightarrow B \rightarrow C$, as displayed.

When no training has occurred all weights are presumed to be zero and no associative retrieval takes place yet. If the element producing a prediction is associatively active (i.e. not actually present, only cued), its prediction is discounted by a factor ϑ. Theta avoids infinite reverberating loops as discussed in (Wagner, 1981, p. 13). At a given time-point, the aggregate of the predictions for an element m of a stimulus o is the total prediction for it, and is denoted $\hat{Y}^t_{m,o}$, and consists of the contribution of each other element of each stimulus active at that time-point, that is the link from the predicting element times its overall activation at that moment in time:

$$\hat{Y}^t_{m,o} = \sum_p \sum_j w^t_{j,p \rightarrow m,o} \cdot \mathscr{A}^t_{j,p} \tag{2.2}$$

Here $\hat{Y}^t_{m,o}$ is the total prediction, and the dot product of weights w and total activities $\mathscr{A}$ is taken over all elements j of all predicting stimuli p to element m of the outcome o. Henceforth j, p will stand for the jth element of predicting stimulus p, while m, o will stand for the mth element of outcome stimulus o.

Due to each element of a stimulus following an approximately Gaussian activation curve, associative predictions from such an element will tend to also follow a roughly similar shape. Thus, when two CSs are paired together simultaneously, they will subsequently predict the elements of the other CS in approximately the temporal order in which they were active directly during a trial (Figure 2.6).

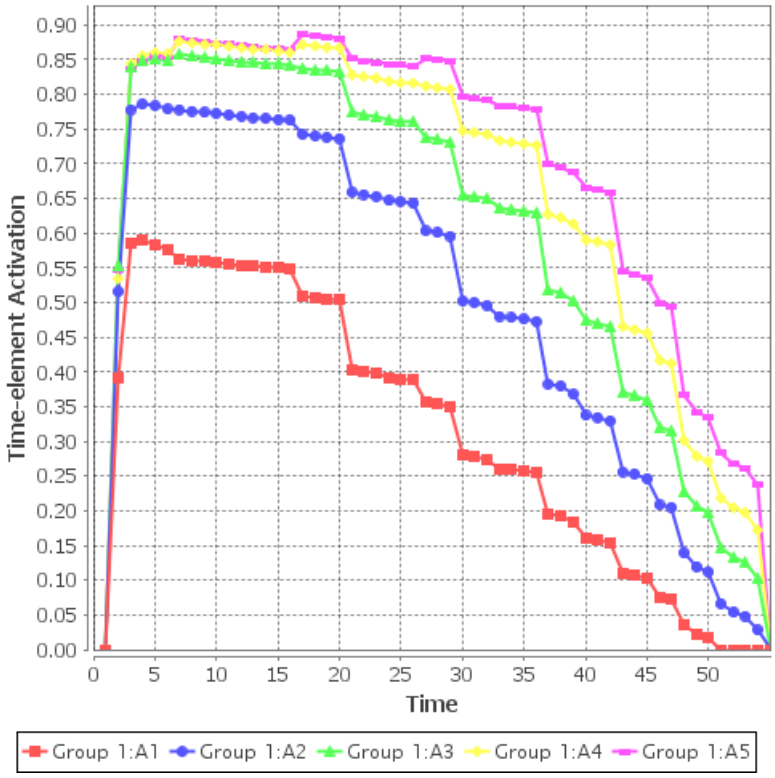


Fig. 2.6 Associative activation of one (absent) CS by another (present) CS, after both have been paired together. Here A1-A5 are predictions for the different time-wave clusters of elements of the predicted CS.

Overall Stimulus Activity

The overall activation of an element m of a stimulus o is taken to be which ever is larger: its direct activation $Y^t_{m,o}$, or its associative activation $\hat{Y}^t_{m,o}$ times the associative discount ϑ:

$$\mathscr{A}^t_{m,o} = max(Y^t_{m,o}, \hat{Y}^t_{m,o} \cdot \vartheta) \tag{2.3}$$

That is, the associative and direct activation of a given element do not compete directly, but rather act in a complimentary fashion in terms of invoking activity of an element. Hence predictions formed by one stimulus for another do not inhibit the activation of the predicted stimulus as in some models such as SOP (Wagner, 1981), but rather simply reduce the novelty of the predicted stimulus. I did so, as when associative and direct activations compete, a CS can for instance reduce its level of responding to an outcome due to it predicting itself or being predicted by the context, which seems to contradict empirical data.

The overall activity of a stimulus is hence the aggregation of the overall activities of the sampled elements of the time-wave clusters. For a standard 10 time-unit CS, the CS activation will (due to probabilistic noise), look as in Figure 2.7.

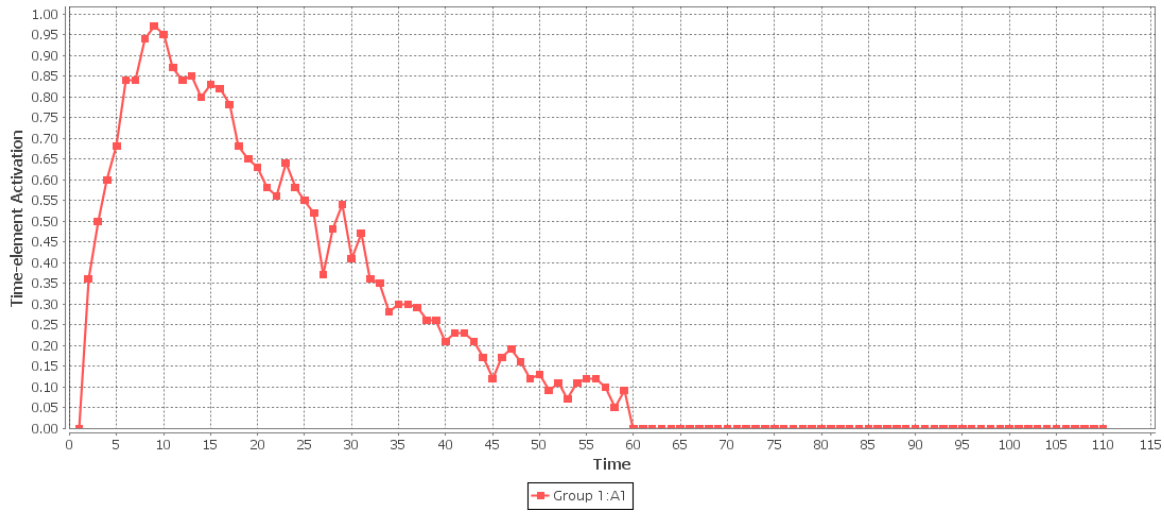


Fig. 2.7 Total activation of a 10 time-unit CS, with persistent memory trace visible. The probabilistically sampled elements of a stimulus aggregate to constitute the overall activation of a stimulus.

Learning and the Dynamic Asymptote

Learning occurs in the model whenever two elements are concurrently active either through sensory experience or associative retrieval. The direction of learning (excitatory or inhibitory) between the elements is dependent on the total predictions made by all cues for a given outcome and the dynamic asymptote of learning used (Figure 2.8), which is a measure of closeness between the activation of the two elements, predicated upon principles of Hebbian learning. That is, if two elements have similar activities they will undergo more rapid conditioning towards one another and eventually reach a higher asymptotic link strength. Hence, two temporally overlapping present stimuli will tend to undergo strong excitatory learning, while a retrieved and a present cue will support a significantly lower maximal level of conditioning. This of course is assuming no cue competition by other cues, as the determinant of the direction of learning is the US error-term, δ_{US}. This mechanism derives from the idea that elements (seen as attributes of stimuli) with a similar activation level are equally probable of being present at that moment, and hence belong in the same 'causal

modality'.

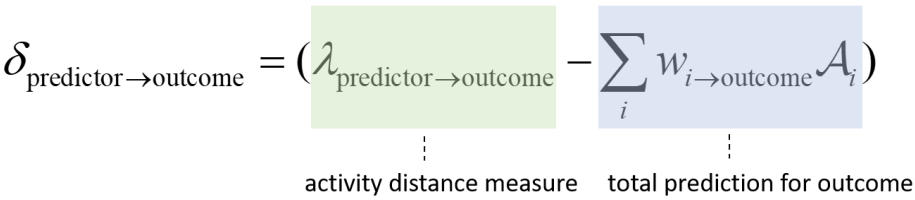


Fig. 2.8 The direction of learning between a predictor and outcome in the DE model is determined by the outcome's error-term, which is influenced both by the similarity in activation of the predictor and outcome, as well as the total prediction for the outcome by all elements.

The power of the dynamic asymptote is that it predicts the same $A1 \rightarrow A1$ and $A2 \rightarrow A2$ learning (though it does not presume SOP activation states) as the Dickinson and Burke SOP learning rules in most cases, while being able to produce $A2 \rightarrow A1$ learning governed by both Dickinson and Burke's and Holland's rules depending on the preceding training (i.e. the A2 stimulus' associative strength towards the A1 stimulus prior to $A2 \rightarrow A1$ conditioning). Yet to be clear, the model does not have simple learning rules; all these effects derive from the effects of the outcome error-term. As the retrieved A2 stimulus will usually have a lower activation than if it was directly present, its dynamic asymptote towards the A1 stimulus will occupy a lower value as compared to $A1 \rightarrow A1$ conditioning. Therefore, its associative strength will approach this intermediate value. If its prior associative strength is lower than this intermediate value (e.g. as in a mediated conditioning design) it will gain associative strength, and if its prior associative strength was higher (e.g. as in a backwards blocking design) it will lose associative strength. Hence, by presuming that the maximum supported extent of learning is in fact dynamic and proportional to the degree to which two stimuli are equi-present either directly or through associative retrieval, various learning phenomena can be explained in a parsimonious and emergent manner. Further, the strength of a retrieved association is crucial for the extent of excitatory learning between a retrieved cue and a present stimulus. Neuroscientific data for stronger retrieved representations tending to be more associable has been gathered by Zeithamova, Dominick, and Preston (Zeithamova

et al., 2012). They found that the degree to which a cue was associatively reactivated was correlated with subsequent performance on a predictive inference task involving the retrieved cue and a present outcome.

The asymptote of learning used in the outcome error-term is a measure of the distance in activity (with direct activation above zero translated to a value of 1), between the predictor element and the predicted element (a small constant is added to each to avoid division by zero). It is in a sense an inverse distance function. It is given by the following equations:

$$\hat{\mathscr{A}}^t_{j,p} = \begin{cases} 1, & if\, Y^t_{j,p} > 0.1 \\ \mathscr{A}^t_{j,p}, & otherwise \end{cases} \tag{2.4}$$

$$\lambda^t_{j,p\rightarrow m,o} = ||\hat{\mathscr{A}}^t_{j,p}\hat{\mathscr{A}}^t_{m,o}|| = \frac{\hat{\mathscr{A}}^t_{m,o} - |\hat{\mathscr{A}}^t_{m,o} - \hat{\mathscr{A}}^t_{j,p}|}{max(\hat{\mathscr{A}}^t_{m,o}, \hat{\mathscr{A}}^t_{j,p})} \tag{2.5}$$

The result is an asymptote based on a linear distance function, with two cues with highly dissimilar activations supporting less learning than if their activations at a given time-point were more similar. As seen in Figure 2.9, the asymptote's values range between -1 and 1. The former is produced when the outcome representation is completely inactive, while the predictor has an activity level of 1. The latter is produced when the outcome and predictor have the same activity level. As the absolute value is subtracted from the outcome's total activity level, this dynamic asymptote is anti-symmetrical; i.e. the outcome activity is more determinant of whether the asymptote is positive or negative.

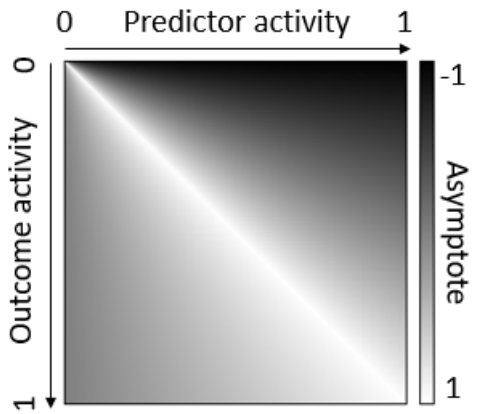


Fig. 2.9 Asymptote of learning as a distance measure of the activities of the predictor and outcome.

In principle many inverse distance functions would fulfil a similar role, however the one used has a number of convenient properties. Firstly, it is linear in that the asymptote between say a present outcome and a retrieved CS scales in proportion to the activity of the CS (presuming the US activity is the same in all cases). One could envision a different inverse distance function which scales say hyperbolically, which would function akin to a receptive field. We however empirically found that using a linear rule seems to produce the most stable and accurate learning. The inverse distance function used is further anti-symmetrical, in that the asymptote of learning from a present cue to an absent outcome is usually negative, and the asymptote of learning from a retrieved (yet absent) cue to a present outcome is high. In this sense it is not a strict inverse distance function like the Euclidean absolute value, or the p-adic distance (see for instance (Koblitz, 1996)), as it violates symmetry and can take on negative values. However the absolute value component of this function is a strict distance function $d : X \times X \to [0, \infty)$, as it meets the following requirements:

1. $|a-b| \geq 0$, $|a-b| = 0$ *iff* $a = b$ (positive measure, identity)
2. $|a-b| = |b-a|$ (symmetry)
3. $|a-c| \geq |a-b| + |a-c|$ (triangle inequality)

The Double Error-Term and Weight Update

The outcome-error (i.e. prediction error) from an element j of a given predictor p to an element m of a given outcome o is given by the discrepancy between the asymptote of learning $\lambda^t_{j,p\rightarrow m,o}$ and the total prediction for the outcome $\hat{Y}^t_{m,o}$:

$$\delta^t_{j,p\rightarrow m,o} = \lambda^t_{j,p\rightarrow m,o} - \hat{Y}^t_{m,o} \tag{2.6}$$

When the element j becomes active after the element m, λ is additionally multiplied by the backward discount b. As the asymptote of the model operates through a distance function of element activities, and these activities persist through time, the model does not utilize the form of gradient learning incorporated by the error-term of the TD model. Links between elements approximate the activation of other elements at a given moment in time instead of predicting the change in these activities.

Uniquely in the model, the predicting element itself has an error term denoting how expected it is, as given by the equation:

$$\delta^t_{j,p} = |\mathscr{A}^t_{j,p} - \hat{Y}^t_{j,p}| \tag{2.7}$$

The reason why the predictor-error is integrated as a part of the learning equation, and not as simply the activity of the predictor (as is done in predictive-coding approaches and some associative learning models), is two-fold. Firstly, this would result in negative activities if the absolute value of the sensory minus predictive inputs is not taken, which seems to deviate from the principle of empirical plausibility. For instance, empirical observation ties dopamine spiking to prediction errors by the presumption that negative prediction errors are

encoded by a lull in dopamine spiking (Schultz et al., 1997). Secondly, we have through experimentation found that even if such an absolute value is taken, it results in predicted cues in return making weaker predictions for other cues. Thus, a good predictor of an important outcome is often unable to predict the outcome strongly due to the context or other CSs predicting it strongly.

These predictor and outcome error-terms are then used in the weight update equation for a given element, along with the salience s of the predicting and predicted element, their respective activations $\mathscr{A}$, as well as the adaptive α_n or α_r (see next section) for a respectively non-reinforced or reinforced outcome:

$$\Delta w^t_{j,p\rightarrow m,o} = \delta^t_{j,p\rightarrow m,o} \cdot \delta^t_{j,p} \cdot s_{m,o} \cdot s_{j,p} \cdot e^t_{j,p\rightarrow m,o} \cdot \mathscr{A}^t_{j,p} \cdot \mathscr{A}^t_{m,o} \cdot \alpha(j,p)^t \tag{2.8}$$

That is, the direction of learning is determined by the outcome prediction-error and its unique asymptote, while the prediction error for the predictor itself, along with other modulating factors (the fixed salience s, a stabilizing eligibility e, the adaptive α, and activations $\mathscr{A}$) influence the extent and speed of learning. As such, the novelty of the predicting element plays a crucial role in determining the speed of learning. That is, more novel cues are more readily associated with an outcome. The source of this novelty, namely the extent to which other cues have formed associative links to the predictor of the outcome, is fully accounted for by the model. This crucially allows the DE model to explain pre-exposure effects such as latent inhibition in a parsimonious manner without postulating ad hoc psychological processes beyond associative learning and attentional processes related to it.

Attentional Modulation

The revaluation alpha (denoted by α_r or α_n) of the model works by the principle that attention to reinforced or non-reinforced cues increases when there is uncertainty in the occurrence of stimuli of that respective class. It further postulates that if the uncertainty remains high over a sufficient period of time (as tracked by the moving average), the animal uses this persistent uncertainty as a source of information, and reduces its level of attention to such classes of cues (Figure 2.10). Hence, when the occurrence of reinforcers is uncertain, the speed of acquisition of excitatory or inhibitory links from elements of any stimulus to elements of a reinforcer increases initially. Likewise, uncertainty in the occurrence of non-reinforcing stimuli similarly increases the learning speed towards them. If this uncertainty however remains high, the level of attention decays again. The change in the attentional modulation is in proportion to the time-dependent activation $\mathscr{A}^t_{j,p}$ of the element as well as ρ, which is an adaptation rate parameter determining how quickly the revaluation alpha changes. In terms of notation r/n and US/CS denote that one or the other is meant. The overall direction towards which the α changes is determined by the overall moving-average error of its respective class of cues, $|\delta^t_{US/CS}|$:

$$\alpha_{r/n}(j,p)^t = (1-\hat{\delta}^t_{US/CS})(1-\rho_{r/n}\mathscr{A}^t_{j,p})\alpha_{r/n}(j,p)^t + \rho_{r/n}\mathscr{A}^t_{j,p}|\delta^t_{US/CS}| \tag{2.9}$$

Here δ is a moving average of the error terms of respectively all active reinforced or non-reinforced elements, and $\hat{\delta}^t_{US/CS}he$ is a decay which kicks in if the moving average crosses a threshold. For instance, during partial reinforcement it leads to the α_r decaying after sufficiently many trials when the animal realises that the contingency is inherently random and therefore warrants less attention. Thus, the crux of the introduced alpha is that an animal can turn a lack of certainty over the occurrence of cues into a source of information in itself. Now, the overall error of its respective class of cues, $|\delta^t_{US/CS}|$ is updated

per the following formula. Here *o* and *m* respectively denote the stimuli and their constituent elements of the given class (reinforced or non-reinforced) of cues over which the errors are aggregated.

$$\delta^t_{US/CS} = (1 - \rho_r \mathscr{A}_{j,p}) \hat{\delta}^t_{US/CS} + \rho_r \mathscr{A}^t_{j,p} \left| \frac{\sum_o \sum_m \delta^t_{m,o}}{\#\ CS + US\ elements} \right| \tag{2.10}$$

$$\hat{\delta}^t_{US/CS} = \begin{cases} \delta^t_{US/CS}, & \text{if } \delta^t_{US/CS} > \xi \\ 0, & \text{otherwise} \end{cases} \tag{2.11}$$

Hence, the DE model assumes that selective attention towards reinforced and non-reinforced outcomes is proportional to the overall uncertainty of these classes of stimuli, which I believe is in concordance with empirical data concerning attention and prediction-error signals, such as (Holland and Schiffino, 2016). This For instance, in a standard acquisition and extinction procedure, the α_r will rise both during the beginning of acquisition and the beginning of extinction (Figure 2.11). It further predicts that changes in associability induced by uncertainty in one outcome will influence the associability of cues towards other outcomes, and likewise for non-reinforced learning. In psychological terms this implies that uncertainty in the occurrence of reinforced or nonreinforced stimuli increases the processing of such classes of cues, and that the relative processing of non-reinforced and reinforced classes of cues depends on the relative uncertainty of reinforcers when compared to non-reinforcing cues. Thus, the effects of extant reinforced learning (i.e. the extent to which a stimulus is predicted by other stimuli) upon the rate of further reinforced learning is mediated both through the outcome error-term being more minute, as well as through the alpha to a reinforcer decreasing proportionally to learning. Interestingly, although these alphas might seem to produce effects opposite of those in the Mackintosh model (due to their similarity to

the PH formulation), the instantiation of this attentional modulation in real time, and with a sufficiently long CS, in fact produces an effect whereby a better predictor of an outcome attains a higher α_r value (Figure 2.12). This occurs due to the better predictor of the outcome producing a larger prediction error for the outcome before the onset of the latter.

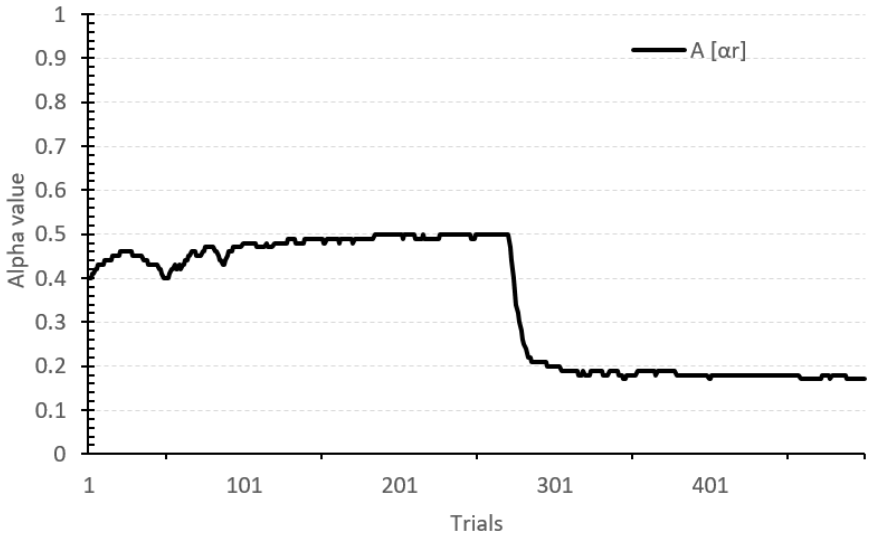


Fig. 2.10 α_r value of cue A in a simulation of partial reinforcement. Once the moving average of α passes a threshold, sufficient evidence has accumulated that the contingency is inherently random, and thus attention decays.

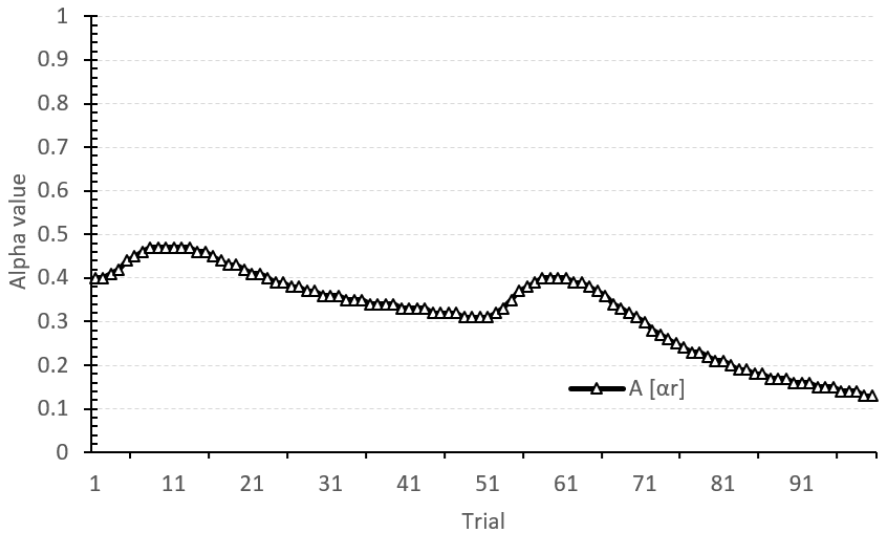


Fig. 2.11 α_r value of cue A in a simulation of 50A+/50A-. The α_r value rises at the beginning of both acquisition and extinction due to the change in contingency producing prediction errors for the outcome.

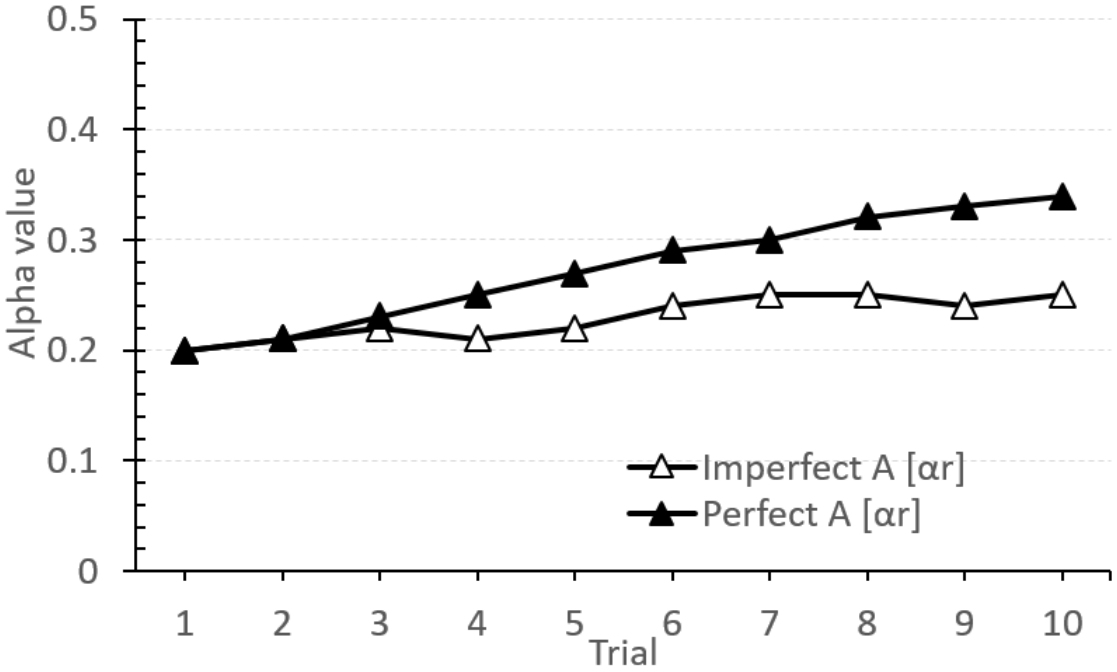


Fig. 2.12 α_r values of 10 time-unit CSs with imperfect (10 intermixed A+ and A- trials) and perfect (10A+ trials) predictors of an outcome. The revaluation alpha rises to a higher level for the perfect predictor due to it producing a larger outcome error before the outcome onset.

Timing and the anticipatory CR

In trial-based error-correction, the asymptotic strength of the link between a CS and outcome will tend towards (given that $\lambda = 1$) a value of 1 with sufficient acquisition training.

However, in the real-time case, one encounters a few obstacles if one aims for the asymptotic strength of an associative link between a CS and US to be able to grow to a value of 1. The first problem is that if one replaces the asymptote with the activity of the US, this will lead to sub-maximal acquisition, as learning both before and after the maximal US activity will deviate from the ideal asymptotic value. Secondly, if one assumes that an anticipatory CR for the US is generated by predictions by the CS proportional to the product of the CS activity and its weight/link to the US, this by proxy entails that a prediction-error of negative sign is generated at each time-point before the US onset. Thus, some extinction will occur on each acquisition trial before the US onset.

There are numerous ways of circumnavigating such difficulties. In the first case, one could specify that the US becomes active as a step-function, such that its activity takes on values of 0 or 1. As this seems to contradict the general assumption we have made of stimuli leaving behind memory traces, we have taken an alternative approach. This approach dictates that the asymptote is given, as mentioned previously, by the similarity in CS and US activity levels, tweaked such that a present CS and US will have an asymptote of learning of 1. With regard to the problem of extinction before the US onset, a possible solution would be to 'discretize' the CS-US learning such that error-correction only takes place at the time of the US. For this to work, one would either need to manually specify to the model when such time occurs (which would violate the principle of using only knowledge available to the animal through known learning processes), or would require a mechanism of timing. As this latter mechanism seems outside the purview of associative learning and thus violates parsimony,

we have instead relied on both the dynamic asymptote and a type of 'eligibility', $e^t_{j,p\to m,o}$. On time-points when the CS is active, when the US has not occurred yet, the prediction of the CS for the US nevertheless activated the US representation in proportion to the link between them. This in return means that the asymptote of learning from the CS to the US is an intermediate value between 0 and 1, thus reducing the amount of extinction occurring. In this manner, the asymptote of a sufficiently short CS will tend to be very close to 1 during acquisition. Further, the eligibility used is calculated as follows:

$$e^t_{j,p\to m,o} = \max(0.1, w^t_{j,p\to m,o} \cdot \mathscr{A}^t_j) \tag{2.12}$$

Hence, time-points on which the predictor predicts the outcome more strongly are more eligible for learning. This tends to occur around the time of the US onset, thereby attenuating the aforesaid extinction of the predictor-outcome link before the outcome onset.

Computation of Trial Values

As a real-time model inherently contains more information about the time-course of predictions and learning between stimuli, one must commit to a fixed mapping from time-based to trial-based values. This is complicated by a few factors. Firstly, different experiments measure trial-based responding using different measures, such as aggregated responses, trials to criterion, responses per time-unit, responding at the time of the US, and maximal responding. To enable the easiest and most consistent methods of comparison to empirical data, we have chosen to calculate trial based weights, associative strengths, and responses by taking the average of these values over the length of the entire trial.

To calculate the response elicited by stimuli on a given trial, the total prediction on a time-point of the trial (equivalent to the associative activation of the US elements divided by their number) is averaged over each time-step of the trial:

$$R^{Trial} = \frac{1}{duration(trial)} \sum_{t=1}^{duration(trial)} \frac{1}{\#elements(US)} \sum_{m=1}^{\#elements(US)} \hat{Y}^t_{m,o} \tag{2.13}$$

Here $\hat{Y}^t_{m,o}$ is the prediction/associative activation of the corresponding US element at time t.

To calculate the associative weight of a stimulus A to another stimulus B, on a given trial, the following formula is used:

$$W^{Trial}_{A\to B} = \frac{1}{duration(A)} \sum_{t=1}^{duration(A)} \sum_{m=1}^{\#elements(A)} \sum_{n=1}^{\#elements(B)} w^t_{m,A\to n,B} \tag{2.14}$$

To calculate the associative strength of a stimulus A to another stimulus B, on a given trial, the following formula is used:

$$V^{Trial}_{A\to B} = \frac{1}{duration(A)} \sum_{t=1}^{duration(A)} \sum_{m=1}^{\#elements(A)} \sum_{n=1}^{\#elements(B)} w^t_{m,A\to n,B} \cdot \mathscr{A}^t_{m,A} \tag{2.15}$$

As can be seen in these equations, the arithmetic mean is taken to calculate the overall stimulus values from the constituent elements. This is done, as the error-correction works on the level of elements. Thus, if a stimulus A has for instance 10 elements, the total associative prediction for these elements could reach a value of 10 if they were simply summed.

Chapter 3

Simulation of Learning Phenomena

3.1 Results of the Double Error Model

To demonstrate the predictive capability of the DE model's unique second error-term, its revaluation alpha, as well as its dynamic asymptote of learning, simulations of phenomena were tested by simulating a set of representative experiments with a general-purpose simulator of the model. Both general simulations (i.e. not replicating a specific, published experiment), and simulations of experiments published in the literature have been conducted, and displayed in each section of this chapter. The use of experimental and simulated controls allows for more thorough testing of the model, as it constrains the explanations possible for each learning phenomenon. The first block of simulations conducted involves simple acquisition and extinction designs, to show how the model's time-dependent processes affect conditioning. Next, the error-correction framework of the model is validated with simulations of conditioned inhibition, blocking, unblocking, and retrospective correlation learning in a learned irrelevance experiment. As it is critical to demonstrate to which extent the DE model's mechanism of a revaluation alpha and learning between motivationally neutral cues can account for pre-exposure related effects (as such non-linear effects cannot be presumed to occur a priori), the third block of results simulated the phenomenon of latent inhibition, a

within-subjects design of perceptual learning, the Hall-Pearce effect, and latent inhibition with compound pre-exposure. Both general and empirical mediated learning experiments (sensory preconditioning, backward sensory preconditioning, backward blocking, unovershadowing, mediated conditioning, and further effects) were then simulated to demonstrate the model's connectionist framework of neutral learning and associative retrieval, along with its dynamic asymptote of learning; and the resultant capability to predict how an animal can re-evaluate past associations through retrieved representations of cues. Next, the competence of the model in coping with compound stimuli and complex, configural discriminations was tested through simulating empirical data on negative patterning and biconditional discriminations, along with simulations testing generalization decrements and a non-linear discrimination involving three stimuli. These simulations are intended to make evident that, with only the assumption of common elements, this elemental model can learn to approximate learning of configurations. Finally, scalar invariance and temporal overshadowing were simulated to show the model can predict some time-dependent effects. Where simulations and empirical data are juxtaposed as curves over multiple data-points, the goodness of fit is approximated using a correlation coefficient R:

$$R(A,B) = \frac{\mathrm{COV}(A,B)}{\sigma_A \sigma_B} \tag{3.1}$$

Here A and B are the empirical and simulated data-sets, COV is the covariance, and σ is the standard deviation.

For each phenomenon, we present both the design and the trial-by-trial response values of interest. Simulations are presented with a response measure matching the experimental output. When required to parallel experimental results, a simulated suppression ratio (r) was

computed following (Mondragón et al., 2014), where:

$$r = \frac{100 - \mathrm{CR}}{(100 - \mathrm{CR}) + 100} \tag{3.2}$$

Here CR denotes the conditioned response calculated for the trial. Thus, values closer to 0.5 indicate poor conditioning whereas values near 0 correspond to high levels of conditioning.

We will follow the order of exposition of the phenomena above. The default parameters of the simulations are displayed in the Table 3.1.

Type		Parameter	Value
Elemental representation		CV	20
		Wave-constant/c	2
		Curve right-skew/s	20
		Set size	10
		Common elements proportion	0.1
Activation discount		Associative activation discount	0.95
Memory discounts		Backward discount/b	0.75
		Eligibility trace discount	0.95
Model		Associability recency	0.01
Associability (α)	α r	CS	0.8
		Ctx.	0.25
	α n	CS	0.4
		Ctx.	0.2
	α+	US	0.2
Salience	s	CS	0.3
	s	Ctx.	0.07
	β	US	0.9

Table 3.1 Parameters for the simulations of learning phenomena.

Here, CV is the coefficient of variation of the curve. The curve right-skew is the degree to which the time-wave is skewed to the right. The set size is the number of elements

in each time-wave cluster. The common elements proportion is the number of elements in common between two CSs. The CV CS element activation is the scalar multiplier of the CS time-wave variance. The associative activation discount multiplies the activity and predictions of all retrieved cues. The backward discount multiplies the asymptote of learning when the time-wave a given predicting element belongs to occurs after the time-wave of the predicted element. The eligibility trace discount multiplies the eligibility value. Finally, the associability recency determines how quickly the moving-average of the revaluation alphas updates at each time-point.

Acquisition and Extinction

Acquisition and extinction of an association will be simulated to demonstrate that the DE model's formulation in terms of its activation function (time-waves) and error-correction produce the expected cornerstone behaviour of a model of learning. Namely that reinforced trials lead to the acquisition of an excitatory CS-US link, and extinction leads to a decay in this link. We will further demonstrate that shorter CSs lead to a higher asymptotic associative weight as expected, as well as various other effects related to simple acquisition and extinction in relation to the model's formulation.

Acquisition, Extinction and CS length

We will begin with a simulation of simple acquisition and extinction of a $CS \rightarrow US$ association, varying the temporal parameters of the model to show how CS, US and ITI timing affect learning in the model. It has been empirically demonstrated that longer CSs support a lower level of acquisition than shorter ones. (Delamater and Holland, 2008) reproduced this finding by studying learning in an appetitive magazine approach task with CSs of lengths

20, 60, and 180 seconds. We will simulate this relation between CS length and the level of acquisition in this section. As is visible in Table 3.2, groups 5CS, 10CS, and 20CS used respectively 5, 10, and 20 time-unit CSs for this simulation. Each group similarly used a 1 time-unit US, and 50 time-unit ITI. All three groups were programmed to undergo 20 acquisition trials followed by 20 extinction trials. The remaining parameters were set per Table 3.1.

Group	Phase
5CS	$20A+/20A-$
10CS	$20A+/20A-$
20secCS	$20A+/20A-$

Table 3.2 Acquisition and extinction design with different CS lengths.

As can be seen in Figure 3.1, acquisition and extinction developed for all simulated CS durations. Of note is that the 5 time-unit CS learns and un-learns the CS-US association faster and to a higher asymptote than the 10 time-unit CS. Similarly, the 10 time-unit CS learns and un-learns the CS-US association faster and to a higher asymptote than the 20 time-unit CS. This effect is more pronounced in extinction, where the timing to the US onset is less relevant.

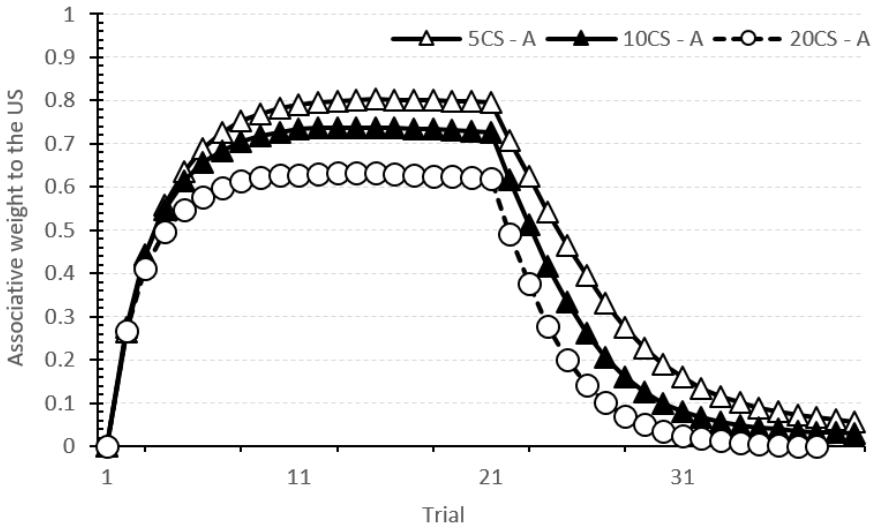


Fig. 3.1 Simulated weights of the CS in the acquisition and extinction procedure with 5, 10, and 20 time-unit CSs.

Forward, Backward, Simultaneous Acquisition

Next, we analyzed the effect of varying the CS onset relative to the US onset. (Rescorla, 1980) has found that simultaneous presentations of two cues can result in strong associations. This is not however always the case, with simultaneous presentations of a CS and US mostly resulting in lower levels of acquisition. Nevertheless, this lower level of responding sometimes observed could be attributable to factors other than the lack of formation of an associative link, such as the use of inappropriate response measures (Arcediano and Miller, 2002), as well as the fact that in simultaneous procedures the CS may in some cases be insufficiently processed before the US onset, and a generalization decrement between the CS-US trials and CS test trials (Nasser and Delamater, 2016). The three conditions considered in this simulation are with the CS terminating at the US onset (Forward), with the US occurring before the CS (Backward), and with the CS onset and US onset co-occurring (Simultaneous). The full design is displayed in Table 3.3. All three conditions use a 5 time-unit CS, 1 time-unit US, and 50 time-unit ITI. The simulation consisted of 20 acquisition trials. The remaining

parameters were set per Table 3.1.

Group	Phase
Forward	$20A \rightarrow +$
Backward	$20+ \rightarrow A$
Simultaneous	$20A+$

Table 3.3 Acquisition and extinction design with different CS lengths.

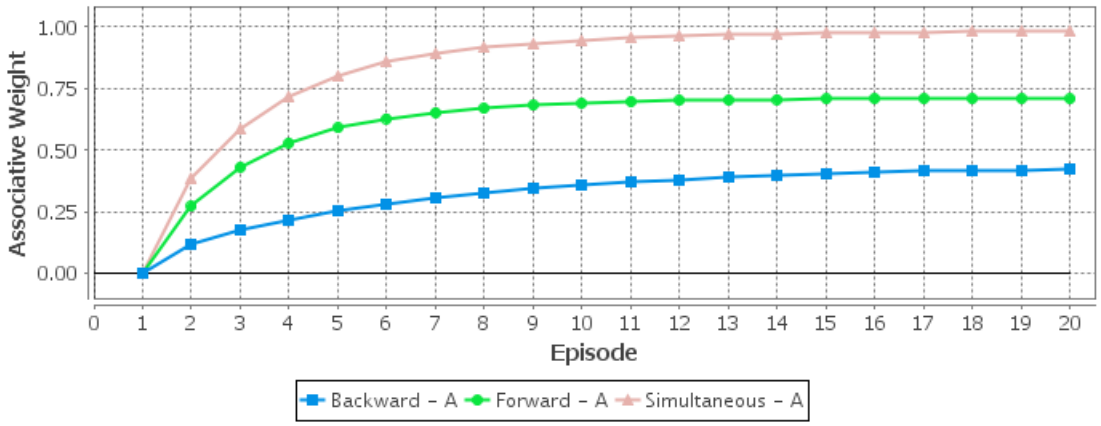


Fig. 3.2 Simulated weights of the CS in the acquisition procedure with forward, backward, and simultaneous conditions.

As is visible in Figure 3.2, acquisition reaches a higher asymptote in the simultaneous timing condition than in the other two conditions, which contradicts most empirical data. This occurs as the temporal overlap between the US and CS is maximized, which due to the dynamic asymptote of the model leads to the maximal amount of learning. This erroneous prediction could be improved by implementing a gradual build-up of the CS and US activities instead of the rapidly growing curves currently used. Similarly the forward condition has a better CS-US activity overlap than the backward condition.

Since parameter b determines the asymptote of backwards learning, setting this value to $b = 0.0$ (i.e. backward excitatory learning is constrained), produces inhibition in the

backward condition (Figure 3.3). What the b parameter denotes is the extent to which strict temporality is effective. That is, a lower b value indicates that co-occurring predictor-outcome activities are more constrained toward supporting temporal maps wherein the cause precedes the effect. Hence, the effect preceding the hypothetical cause must by implication indicate that the hypothetical cause is not causative of the effect, and therefore acquires a negative association to the effect.

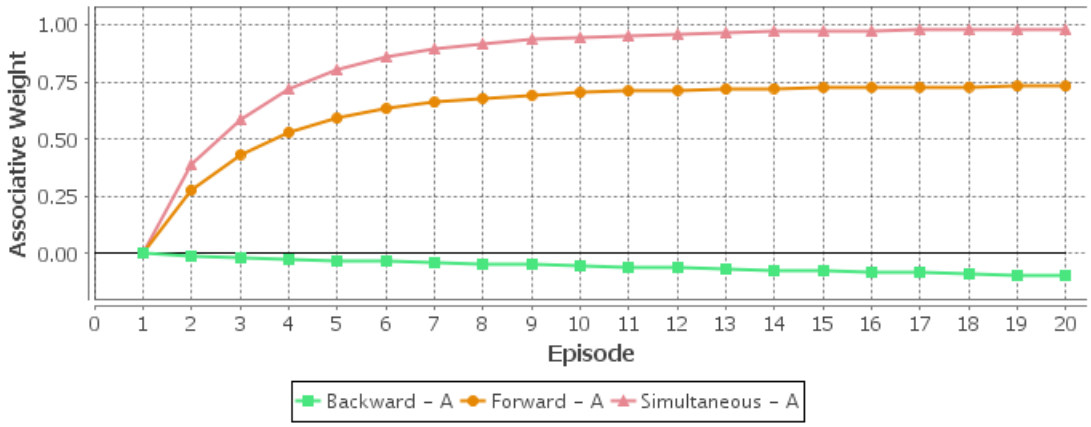


Fig. 3.3 Simulated weights in the acquisition procedure with forward, backward, and simultaneous conditions with $b = 0.0$.

Acquisition & ITI length

A central finding of the field of timing is the phenomenon of the I/T (ITI to Trial) ratio, wherein increasing the length of the ITI in relation to the trial leads to better conditioning. Further, similar I/T ratios have been found to result in roughly similar levels of conditioning (Gallistel and Gibbon, 2000). For the next simulation, we will show that the DE model can reproduce the former result, i.e. how lengthening the ITI leads to stronger conditioning. We simulated acquisition with 10, 30, 60, 120, and 240 time-unit ITIs and a CS of duration 5 time-units, simulated as per the design in Table 3.4. The remaining parameters were set per Table 3.1. Each group underwent 20 reinforced presentations of cue A.

Group	Phase
5sec10sec	20A+ 10sec ITI
5sec30sec	20A+ 30sec ITI
5sec60sec	20A+ 60sec ITI
5sec120sec	20A+ 120sec ITI
5sec240sec	20A+ 240sec ITI

Table 3.4 Acquisition simulation design with different ITI lengths and 5 time-unit CS.

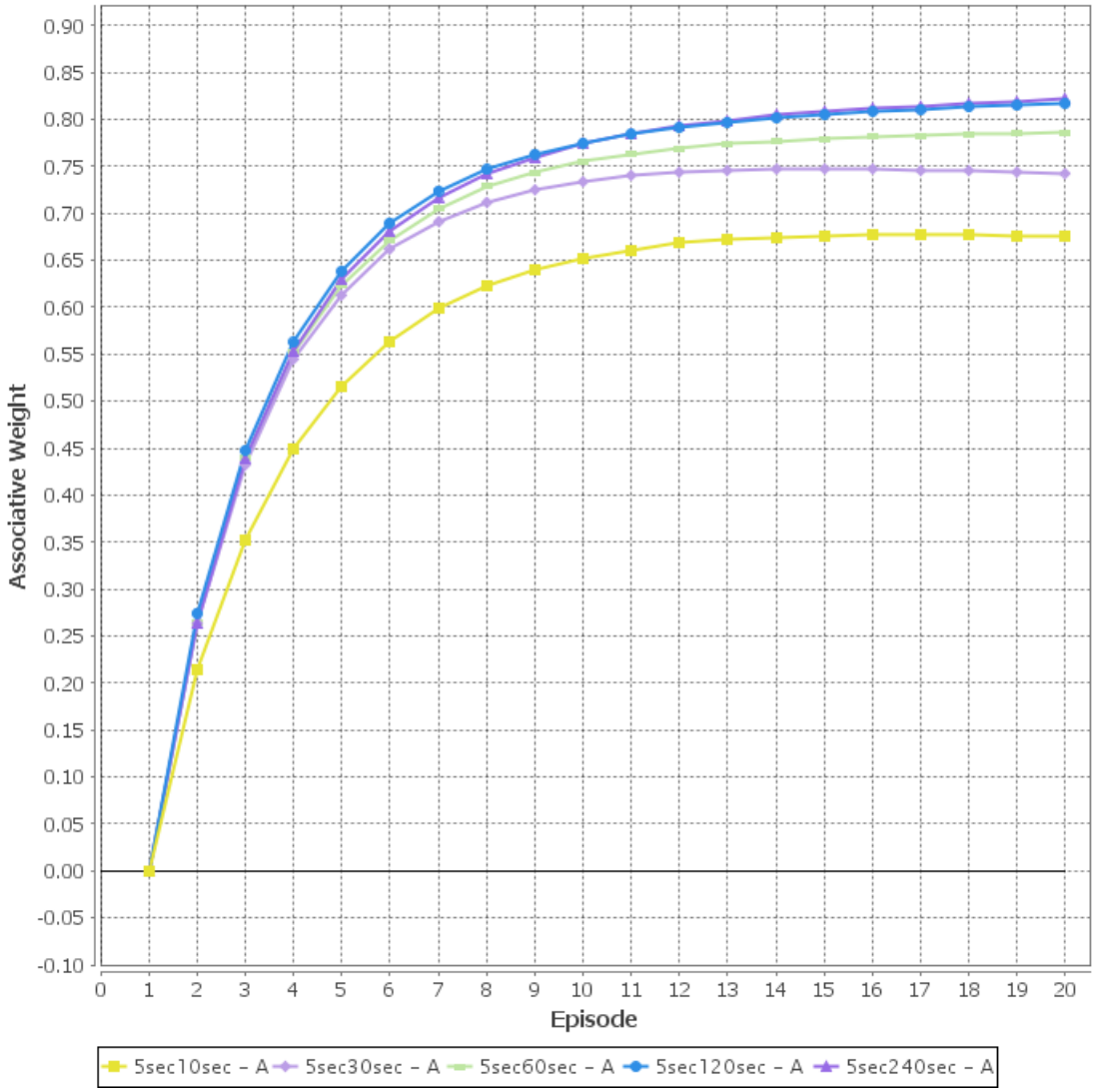


Fig. 3.4 Simulated CS weights of the CS in the acquisition procedure 10,30,60,120,240 time-unit ITIs and 5 time-unit CS.

As is visible in Figure 3.4, longer ITI lengths lead to higher asymptotic acquisition strengths of the CS-US associations, except in the case of 10 second and 30 time-unit ITIs. This was produced due to the context extinguishing its association to the US more effectively with longer ITIs (Figure 3.5), with the difference between the 10 and 30 time-unit conditions probably stemming from the CS overshadowing the context.

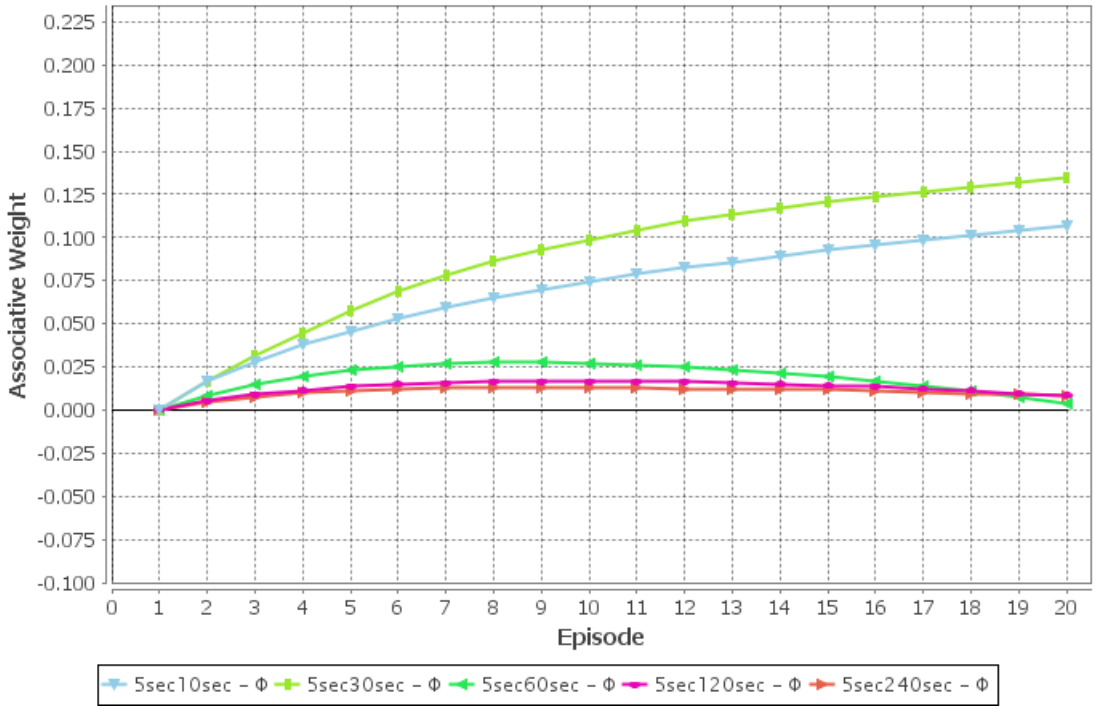


Fig. 3.5 Simulated context weights in the acquisition procedure with 10, 30, 60, 120, 240 time-unit ITIs and 5 time-unit CS. Longer ITIs lead to context extinction.

Activation curve shape & Acquisition

Finally, we will demonstrate the effect of the wave-constant parameter of the model upon the shape of the anticipatory CR in an acquisition procedure. The wave constant is a parameter, which as mentioned contributes towards controlling the standard deviation of the semi-Gaussian activation 'time-waves' of the model, such that higher values correspond to wider curves. In practice, the wave-constant controls the degree of temporal generalization.

This temporal generalization can further interact with the response measure used in an experiment. For instance, if maximal responding during a trial is measured, lower temporal generalization will produce a higher CR spike, which in return will lead to a higher recorded CR. We have simulated 10 acquisition trials with a 10 time-unit CS with wave constant = 1.5 and wave constant = 15. The remaining parameters were set per Table 3.1. The effect of this upon the shape of the time-waves is visible in Figure 3.6. The effect of this upon the real-time responses is visible in Figure 3.7. As can be seen, the simulation with a smaller value for the wave-constant produces an anticipatory CR which grows exponentially, while the simulation with a higher value produces a much more steady CR. Evidence exists for CR timing curves similar to both of these curves, and depends on the exact measure of responses used (Gallistel and Gibbon, 2000).

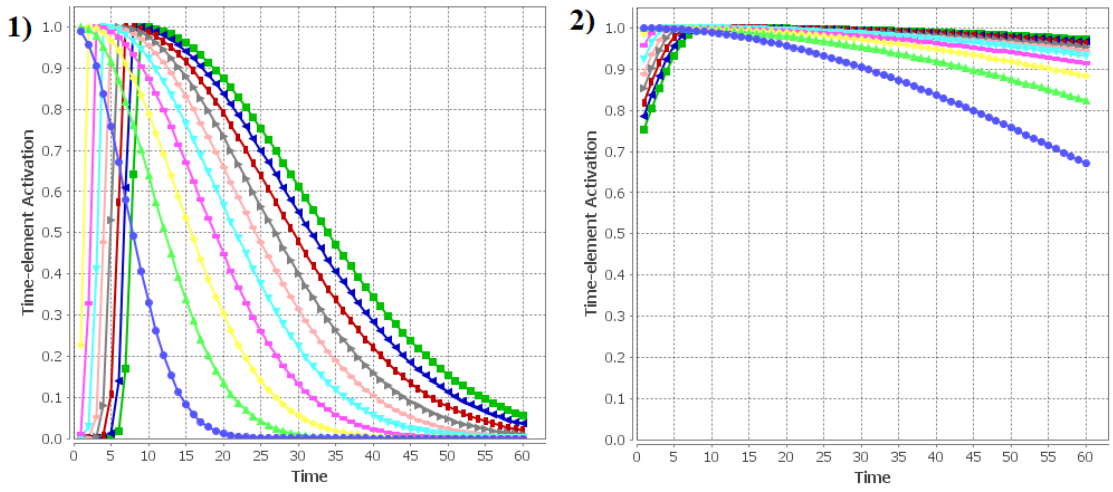


Fig. 3.6 1) Time-waves of a 10 time-unit CS with a wave-constant of 1.50. 2) Time-waves of a 10 time-unit CS with a wave-constant of 15.00.

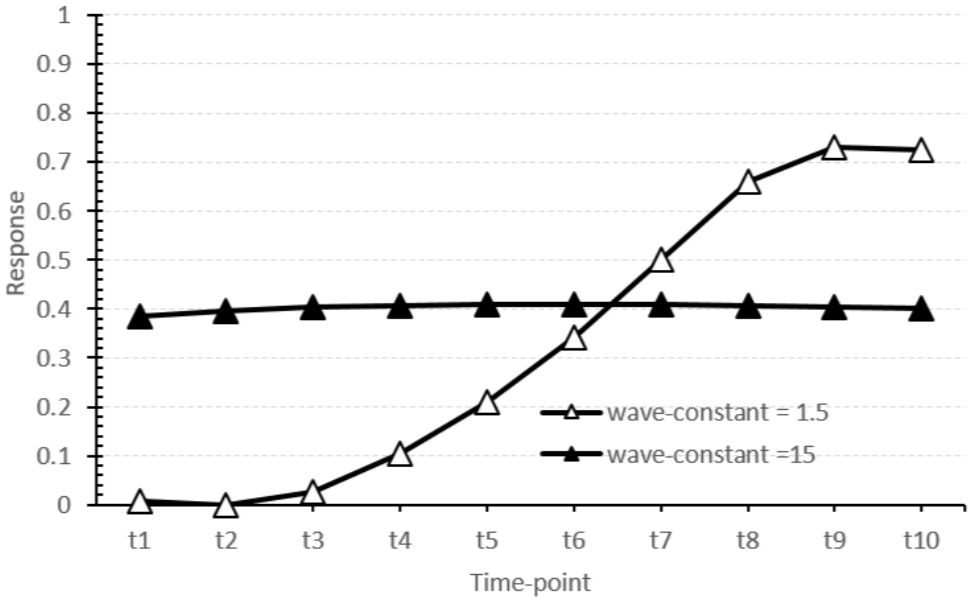


Fig. 3.7 Anticipatory CR after 10 trials of acquisition of a 10 time-unit CS with a wave-constant of 1.50 and 15.00.

Partial Reinforcement

Partial reinforcement occurs when a CS is only followed by a US with a certain probability. Under these conditions, it is expected that the CS will come to predict the US in proportion to the probability of reinforcement, that is to the underlying contingency (Murphy and Baker, 2004; Rescorla, 1968). Reproducing such a result with a model of learning might be trivial, yet many models in fact do not reproduce it. For instance, the PH model does not (Figure 3.8).

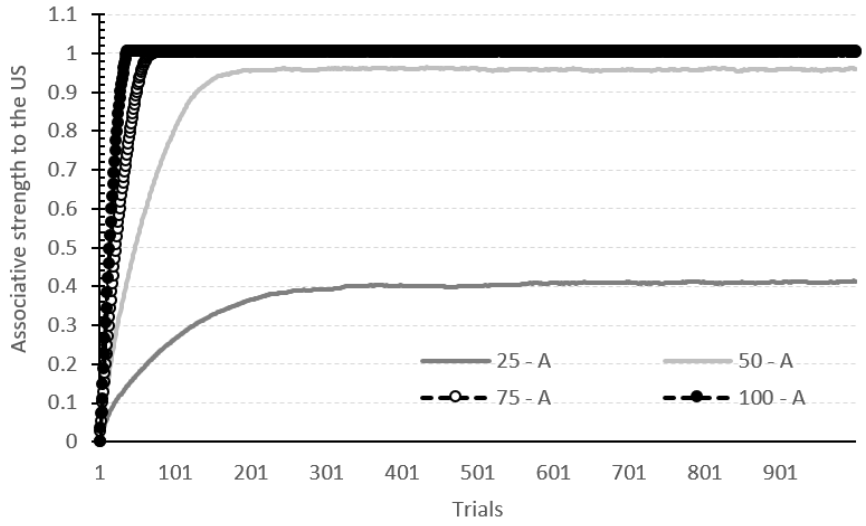


Fig. 3.8 Simulation of partial reinforcement with the PH model using the Pearce-Hall Simultator (Grikietis, R., Mondragón, E., and Alonso, 2016), with reinforcement ratios of 0.25, 0.5, 0.75, and 1.0. The model is unable to reproduce the effect whereby the asymptote of partial reinforcement is proportional to the ratio of reinforced to non-reinforced trials.

To simulate this phenomenon, we have used a design in which the same CS A is followed by random reinforcement with probabilities of 0.25, 0.5, 0.75, and 1.0 in respectively groups 25, 50, 75, and 100. Each group received a total of 1000 A trials, of which the aforementioned ratio were reinforced. The full design of the simulation is displayed in Table 3.5. The simulation was conducted with a CS length of 5 time-units, US length of 1 time-unit, and an ITI of 50 time-units. The remaining parameters were set per Table 3.1. Critical in predicting this aforementioned tendency of partial reinforcement to elicit responding in proportion to the ratio of reinforcement is the model's α_r parameter, which has a built-in decay when the reinforced outcome error remains persistently high, such as in partial reinforcement. This stabilizes the learning after the animal realizes that it is dealing with an inherently random contingency. Figure 3.9, the associative weight of A per group asymptotes approximately in proportion to the ratio of reinforced to non-reinforced trials. However, this is not normative. Contextual units are not controlled to produce this pattern of behaviour; rather, our intention is to show that weights adjust roughly proportionally to the ratio of reinforced to

non-reinforced trials.

Group	Phase
25	$250A+/750A-$ [random]
50	$500A+/500A-$ [random]
75	$750A+/250A-$ [random]
95	$950A+/50A-$

Table 3.5 Partial reinforcement simulation design.

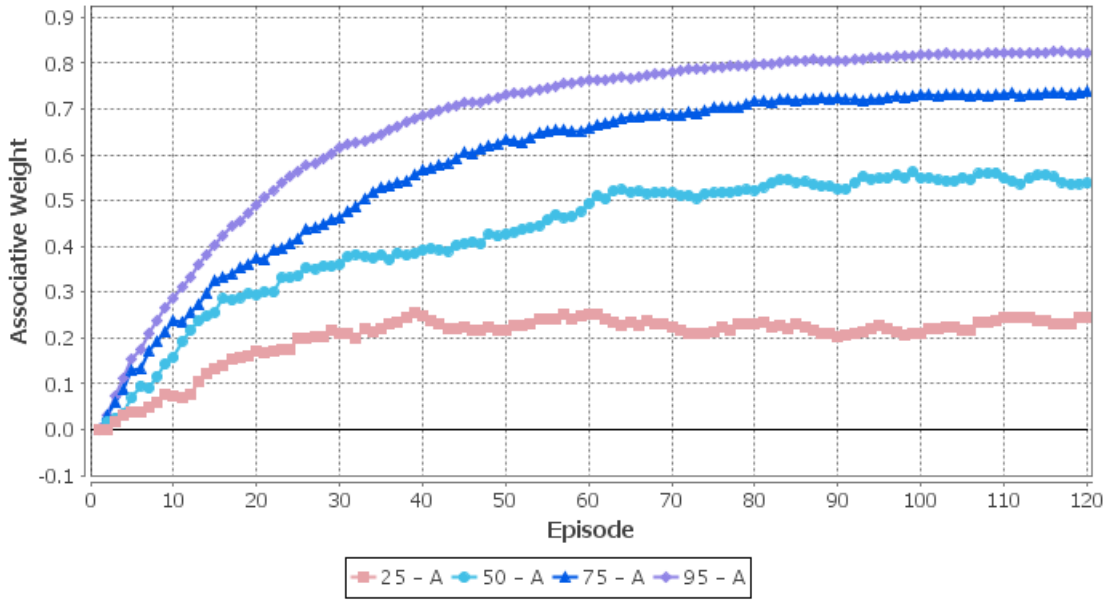


Fig. 3.9 Simulated weights of stimulus A in the partial reinforcement design.

Cue Competition

Cue competition refers to effects wherein the presence of multiple cues on reinforced trials limits the acquisition or expression of learning to each individually present cue, as compared of if such a cue were presented individually. Such competition, following competitive error-correction rules (e.g. the RW model), is thought to manifest more generally when-ever multiple cues are presented. As the DE model utilizes a competitive error-correction rule, we have interpreted effects such as conditioned inhibition, blocking, unblocking, and retrospective correlation learning in learned irrelevance as arising from the competition for US processing between CSs, i.e. cue-competition within error-correction. Of course, different theories of conditioning often favour alternative interpretations, such as attentional effects being determinant.

Conditioned Inhibition

Conditioned inhibition is produced when a stimulus comes to predict that an outcome will not occur (Rescorla, 1969). This is thought to happen, according to error-correction theory, when an expectation of the US occurring (due to being predicted by a stimulus previously paired with the US) is violated, with the conditioned inhibitor compensating by learning to predict the lack of occurrence of the outcome. We favour this interpretation as conditioned inhibition has been found to be US-specific in both forward and backward learning procedures (Kruse et al., 1983). As inhibitory links do not invoke a conditioned response (CR) in most procedures, this form of learning is measured empirically by pairing the suspected conditioned inhibitor stimulus with a stimulus known to have an excitatory association to the US and measuring the resultant decrement in expected responding (summation testing), as well as pairing the expected conditioned inhibitor with a US and measuring the resultant slower acquisition (retardation test). As we have direct access to the DE model's associative

weights, no such summation test is necessary.

We have simulated a conditioned inhibition design that demonstrates the effect of presenting a block of reinforced A trials followed by a non-reinforced set of AB trials, with the equivalent intermixed treatment. We have not used a group with random presentations, as we wish the interpret the US error-term per trial. The design is visible in Table 3.6. The temporal parameters of the simulations used a 5 time-unit CS, 1 time-unit US, and 50 time-unit ITI. Other variable parameters were set as per Table 3.1. Group CI and Group CI intermixed both received 10 reinforced A trials and 10 nonreinforced AB trials. In the case of Group CI intermixed, these trials were intermixed.

Group	Phase
CI	$10A+/10AB-$
CI intermixed	$10A+/10AB-$[intermixed]

Table 3.6 Conditioned inhibition simulation design.

In Figure 3.10, it is evident that the intermixed presentation of $A+$ and $AB-$ leads to less conditioned inhibition than in the block presentation. This is due to the US error staying lower in the intermixed group (Figure 3.11) due to extinction of the $A \rightarrow US$ association on non-reinforced trials bounding the maximal prediction error, and hence the potential of B for becoming a conditioned inhibitor. That is, the periodic extinction of the excitor in the intermixed condition leads to cue B becoming less of a conditioned inhibitor.

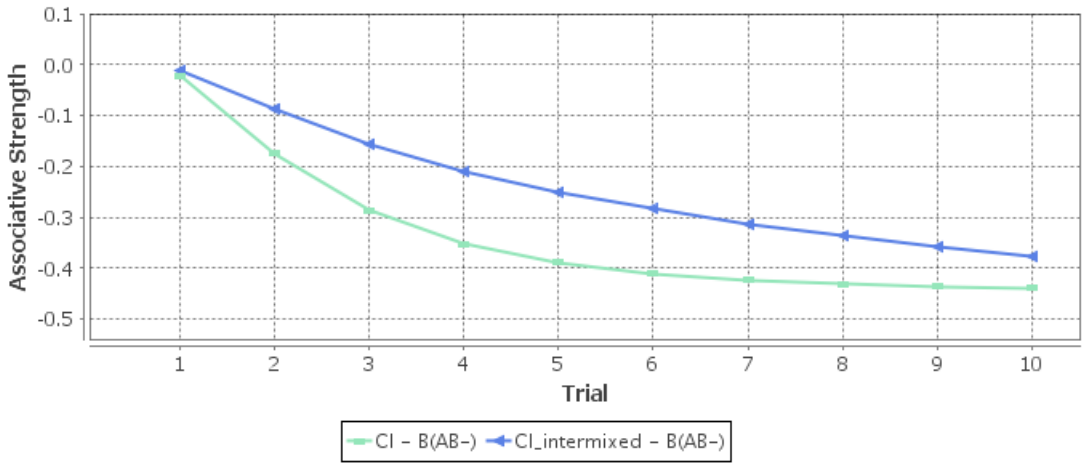


Fig. 3.10 Simulated weights of stimulus B in the CI procedure for block and intermixed presentations.

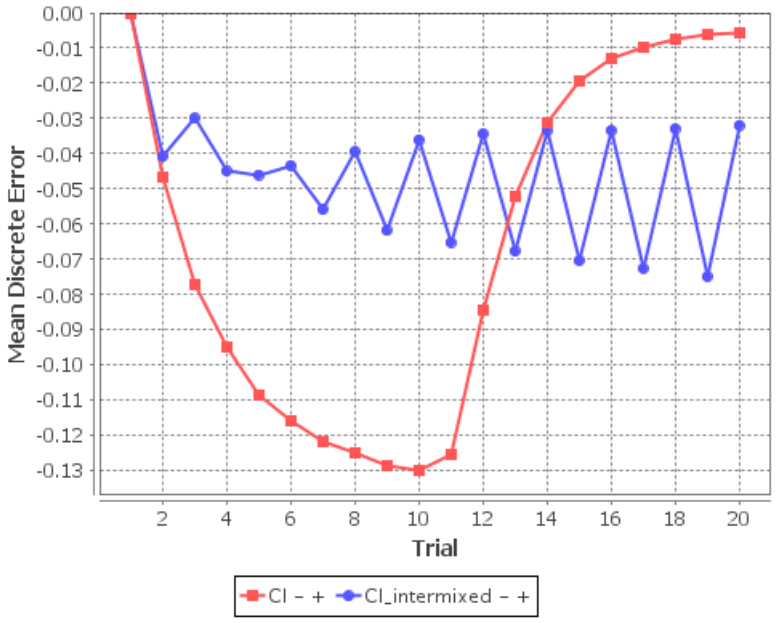


Fig. 3.11 Simulated prediction errors of the US in the CI procedure for block and intermixed presentations.

Blocking

A simulation of the results of a blocking experiment was performed next. Blocking occurs when prior conditioning of a stimulus A with an outcome comes to attenuate subsequent learning between a novel cue B and the outcome in subsequent trials wherein both A and B are reinforced, as compared to a control in which the prior conditioning of stimulus A is omitted (Arcediano et al., 1997; Kamin, 1969).

We have tested blocking with the following design. A blocking and control group were used. Phase 1 consisted of 10 conditioning trials to a 10s stimulus A, ($A+$) for the blocking group. In the control, 10s stimulus N was presented instead ($N+$); In both groups, Phase 2 comprised 10 simultaneous presentations of a 10s stimulus B and the stimulus A ($AB+$). Table 3.7 shows the design of the simulated experiment, while Figure 3.12 displays results (Phase 2) for this experiment.

The parameters used in the simulations are displayed in Table 3.1, except no common elements were used. A CS duration of 5 time-units, US duration of 1 time-unit, and ITI duration of 50 time-units were used.

Group	Phase 1	Phase 2
Blocking	$10A+$	$10AB+$
Control	$10N+$	$10AB+$

Table 3.7 Blocking design.

Figure 3.12 and following figures are saved outputs of the simulator. Depicted in Figure 3.12 is the attenuated conditioning to a blocked stimulus B in comparison to stimulus B in the control group.

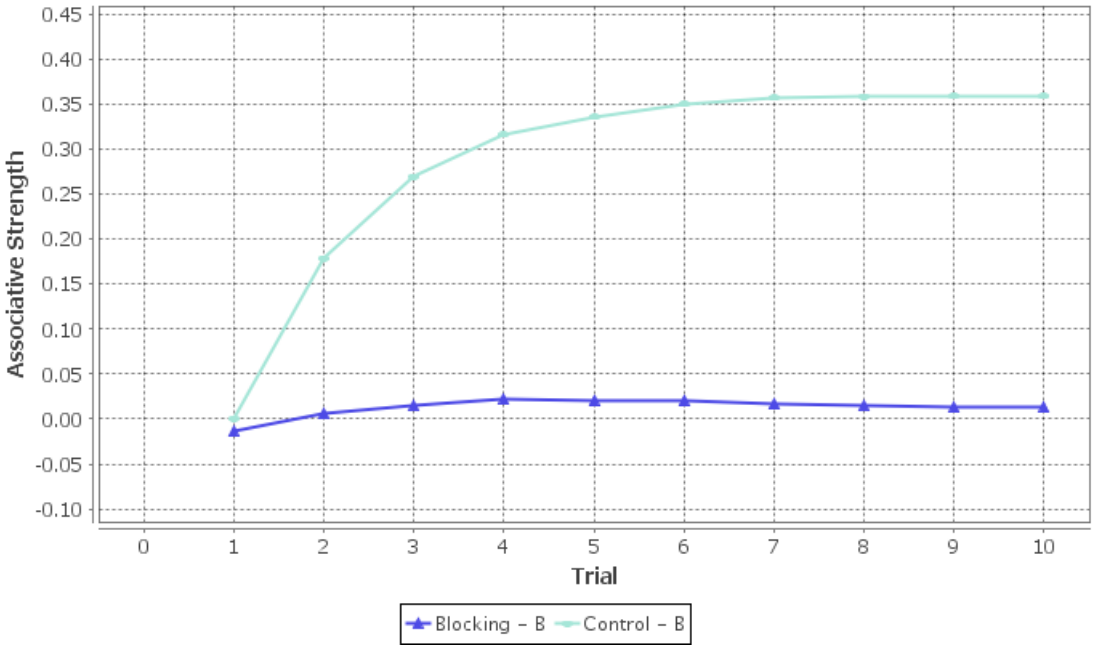


Fig. 3.12 Simulated weights in the second phase of the blocking procedure.

The conditioning curve for the blocking group grows slower and reaches an earlier asymptote than the control group (Figure 3.12). The blocking stimulus A can be thought to reduce the associability of the blocked stimulus B, by reducing the overall error term δ_{US} of the US (Figure 3.13). That is, stimulus A gains excitatory strength during reinforcement, which in the subsequent compound reinforcement trials lead to a lower unexpectedness for the US. This in return leads to attenuated learning between blocked stimulus B and the US.

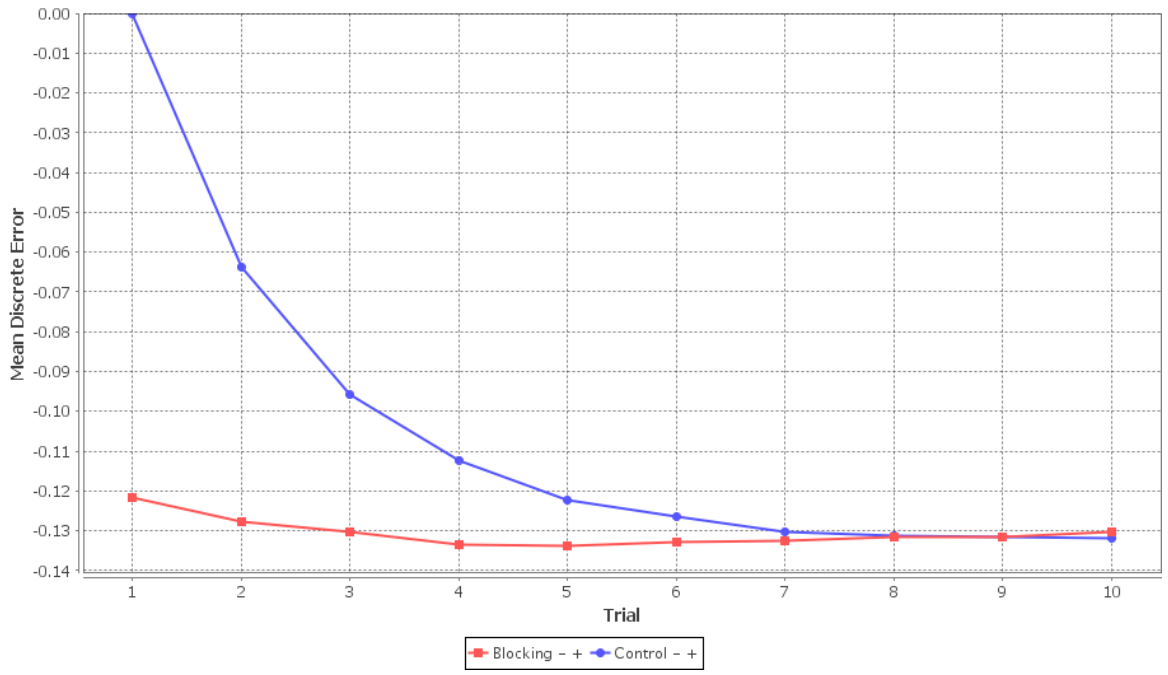


Fig. 3.13 Trial-by-trial prediction error for the US in the blocking simulation.

Unblocking

In Experiment 1 (Bradfield and McNally, 2008), three groups of rat received reinforced presentations of a CS A. In the second phase of the experiment, all three groups received reinforced presentations of CS compounds AB and CD. In Group Block the same reinforcement was used as in the previous phase. In Group Unblock intensity, a higher intensity of reinforcement than in stage 1 was used. In Group Unblock number, the compounds were reinforced by two serially presented USs. The experiment found (Figure 3.14) that both an increase in the US intensity (Group Unblock intensity) and an increase in the number of US reinforcements (Group Unblock number) produced less blocking, i.e. unblocking, as assessed by the test phase.

Group	Phase 1	Phase 2	Test
Block	$A+$	$AB+/CD+$	$B?/D?$
Unblock intensity	$A+$	$AB+/CD+$	$B?/D?$
Unblock number	$A+$	$AB++/CD++$	$B?/D?$

Table 3.8 Design of experiment 1 of (Bradfield & McNally, 2008).

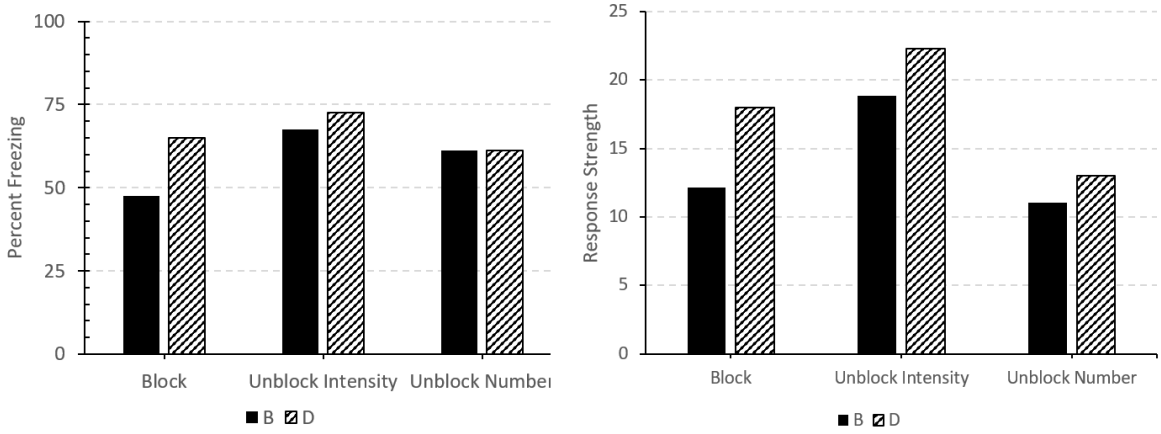


Fig. 3.14 Left: Mean percent freezing in the test phase of experiment 1 of Bradfield & McNally (empirical units adapted from paper), Right: corresponding response strength in the test phase in simulation of Experiment 1.

The simulation was conducted with a 1 time-unit CS, 1 time-unit US, and 50 time-unit ITI. The starting values for the CS α_r and α_n parameters were respectively 0.4 and 0.8. The CS and context saliences used were respectively 0.1 and 0.01. The stimulus intensity in phase 1 was 1.0 for all groups, and 1.5 for Group Unblock intensity in Phase 2. The remaining parameters were set as per Table 3.1. The simulation (Figure 3.14) reproduced the empirical unblocking results for Group Block and Group Unblock Intensity, that is (in the latter) an attenuation of the suppression of responses to the stimulus paired with a novel CS in relation to that paired to the CS conditioned in Phase 1. It further reproduced the direction, but not magnitude, of the unblocking result in Group Unblock Number. Both Group Unblock intensity and Group Unblock number displayed a smaller variance between the responses generated by cue B and D in the test phase as compared to Group Block, mirroring the results of the experiment. The model predicts this result in a straightforward fashion due to the increased US intensity in Group Unblock intensity leading to a higher supported asymptote of learning for the US. That is, the maximal strength of the CS to US link increases in the second phase in Group Unblock intensity, thereby increasing the US uncertainty and error term. This increased error-term then allows for CS B to acquire further strength. In Group Unblock number, the result is more subtle. As we have encoded the increased presence of the US as the US being active on more time-points, this decreases the blocking of cue *B* by increasing the temporal activity of the activity of this cue and the outcome. Hence this simulation makes the implicit assumption that animals process multiple US presentations as repetitions of the activity of a single US, with no qualitative differences. An improved result for the Unblock number group could be obtained by changing the parameters of the model, however we have avoided doing so to keep the parameters as consistent as possible across simulations. The importance of this experiment is that a similar result is predicted by the RW model through its fixed λ being higher in the intensity group; however the DE model uses a varying asymptote, and as such this result demonstrates the predictive capability of this

variable asymptote.

Though the Double Error model can predict unblocking through increased US intensity, the model is incapable of predicting unblocking through reinforcer omission, as seen for instance in (Dickinson et al., 1976). This is of note, as it is predicted by the selective attention used in the Mackintosh model.

Learned Irrelevance and Retrospective Negative Correlations

So far, we have seen effects that can potentially be ascribed to a simple loss in the associability of a stimulus. In a learned irrelevance procedure however the uncorrelated presentation of a CS and a US results in an attenuation of the rate of conditioning that is larger than the delay resulting from the cumulative effect of each exposed stimulus by itself ((Bonardi and Hall, 1996; Mackintosh, 1973), etc.). Unlike the phenomena previously described, this effect seems to imply that animals learn that two events can in fact be negatively correlated, i.e., that a stimulus predicts a reduction in the probability of the occurrence of another one. Moreover, evidence has been produced that such learning could also appear in schedules that may require revaluating retrospectively the correlation between the CS and the US, producing inhibitory learning from the CS to the US. In this section, we simulate an experiment showing exactly such retrospective revaluation of inhibition.

In (Baker et al., 2003) Experiment 1, the effect of uncorrelated CS-US presentations upon subsequent acquisition was studied. Specifically, the experiment sought to uncover whether a CS presented in an uncorrelated fashion with an outcome would become a conditioned inhibitor of the outcome as compared to a control in which stimuli were presented separately in two blocks of trials. The authors argue that this control, used in other experiments that

failed to obtain support for uncorrelated learning, could in fact be inadequate because it might also promote uncorrelated learning, thus rendering the CS inhibitory.

The experiment was split into two sub-experiments. Experiment 1a was designed to assess inhibitory properties by means of a retardation test. Experiment 1b was intended to evaluate inhibition with a summation test. During Phase 1 of Experiment 1a animals did not receive any treatment. In Phase 2, Group N/Sh received uncorrelated presentations of a noise and a shock. In Group N+Sh a block of noise trials followed by a block of shock trials were scheduled. In Phase 3, both groups received conditioning trials to the noise followed by the shock (retardation test). In Experiment 1b a light was conditioned to the shock in Phase 1. Group N/Sh and Group N+Sh received identical treatment as their counterparts in Experiment 1a. In Phase 3, all animals were given non-reinforced trials with the noise and a compound noise-light (summation test). During Phase 4 a saving test consisting of reinforced noise trials intermixed with the compound noise-light was carried out. The retardation test in Experiment 1a showed that uncorrelated presentations of the noise and the outcome (Group N/Sh) produced a smaller deficit in subsequent acquisition than the scheduled presentations in Group N+Sh did. The left panel of Figure 3.16 shows the retardation test results in Experiment 1a: Group N/Sh displayed a lower suppression ratio (and hence faster learning) than Group N+Sh.

Figure 3.17 toptom panel shows the results of the summation test. Animals trained under a blocked schedule of presentations (Group N+Sh) showed a stronger summation effect, visible as a larger differential responding to the stimulus and the compound. The summation effect was also evident, but to a lesser degree for animals trained in the uncorrelated treatment of Group N/Sh. These results suggest the formation of an inhibitory association between the noise and the shock, inhibition that was stronger in Group N+Sh.

Group	Ph. 1	Ph. 2	Ph. 3	Phase 4
N+Sh(1A)		$40N-/40+/8N+$ [rand.]	$36N+$	
N/Sh(1A)		$48N-/48+$ [inter.]	$36N+$	
N+Sh(1B)	$8L+$	$40N-/40+/8N+$ [rand.]	$4NL-/2L-$	$40NL-/20L+$ [inter.]
N/Sh(1B)	$8L+$	$48N-/48+$ [inter.]	$4NL-/2L-$	$40NL-/20L+$ [inter.]

Fig. 3.15 Experiment 1 design of learned irrelevance in Baker et. al.

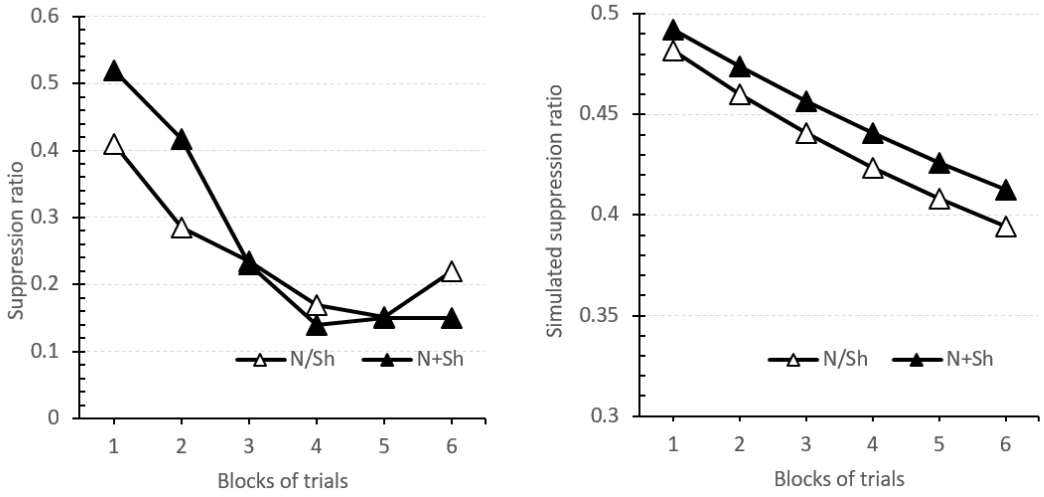


Fig. 3.16 Left: Suppression ratios in experiment 1A of Baker et al. 2003 (empirical units adapted from the paper) for the acquisition test trials in the third phase, Right: corresponding simulated suppression ratios for the same trials. Average correlation $R = 0.92$.

For the simulation, temporal parameters consisted of a 5 time-unit CS, 1 time-unit US, and 50 time-unit ITI. The CS α_n starting value was 0.8, and the CS salience 0.1. The remaining parameters were set to the values in Table 3.1. For Experiment 1a no event was programmed to occur in Phase 1. In Phase 2, Group N/Sh received random presentations of N and of the US, such that both stimuli coincided on 8 trials. Group N+Sh received a block of 48 N trials followed by a block of 48 US trials. In Phase 3 both groups received 36 reinforced N presentations. In Experiment 1b, Phase 1 consisted of 8 reinforced L trials. Phase 2 was programmed identically to that in Experiment 1a. In Phase 3, non-reinforced randomly presented trials of a compound NL and two of L were programmed in both groups. Finally, Phase 4 was programmed to deliver 40 non-reinforced NL and 20 reinforced L trials

randomly presented. The full experimental design is depicted in Table 3.15.

The simulation results match empirical data. In the retardation test (Figure 3.16, right panel), Group N+Sh showed less suppression of responding than Group N/Sh. Results for the summation test (Figure 3.17, bottom panel) closely matched the experimental results. Group N+Sh showed a larger difference in simulated suppression between the NL and L trials than Group N/Sh, suggesting that the noise had become a stronger conditioned inhibitor of the outcome in Group N+Sh.

The DE model accounts for this result by assuming that an animal learns to approximate the CS-US relationship, which in the case of the N+Sh groups would imply retrospectively assessing the information in the following manner: in this group, there will be many US trials in which N is predicted by the context, yet is not present. Thus, the context associatively retrieves N but its activation is weak. The discrepancy between the activation levels of N and the US would yield a minute or negative asymptote, engendering inhibitory learning between N and the US, which combined with competition by the context would allow the model to replicate the result. In contrast, N in N/Sh groups sometimes coincides with the US, therefore forming a slightly less inhibitory link toward it (Figure 3.18). Additionally, the difference between the degree of inhibition attained in each group is further enlarged by the fact that in N/Sh groups the random ordering of the trials bounds the maximal extent to which the context predicts the US in comparison to the prediction following the blocked presentations in N+Sh groups, thus limiting context competition and therefore acquisition of inhibition. In other words, a further source of difference between the groups is due to context-mediated inhibitory learning between the CS and the US.

Of note is that as a consequence, with the parameters used, the DE model does not predict the learned irrelevance result predicted by the selective attention of for instance the

Mackintosh model. This is of note as a similar issue was noted for the model not accounting for unblocking through reinforcer omission, another result predicted by the Mackintosh model.

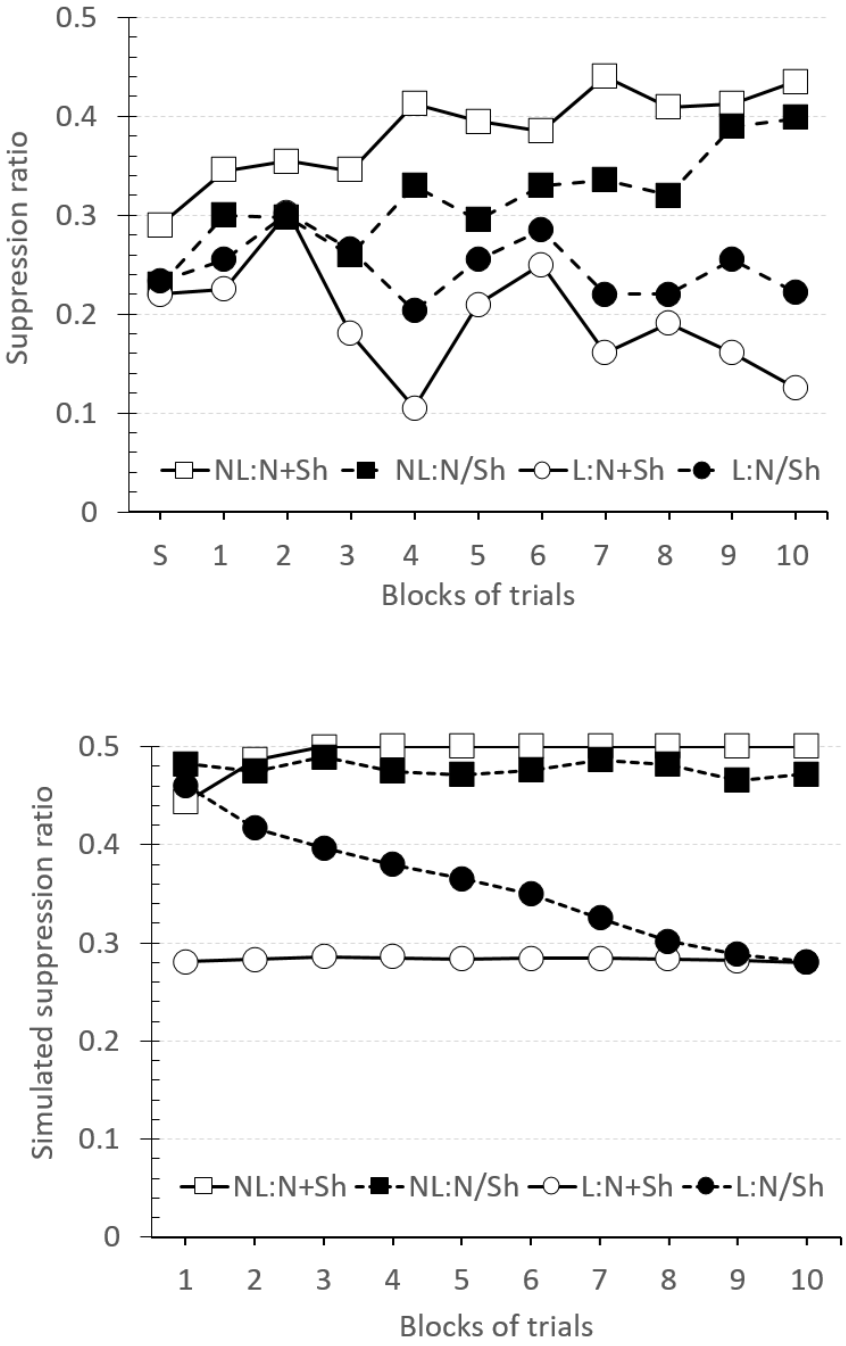


Fig. 3.17 Top: Suppression ratios in experiment 1B of Baker et al. 2003 (empirical units adapted from the paper) for the retardation test trials in the third phase, Bottom: corresponding simulated suppression ratios for the same trials. Average correlation $R = 0.54$.

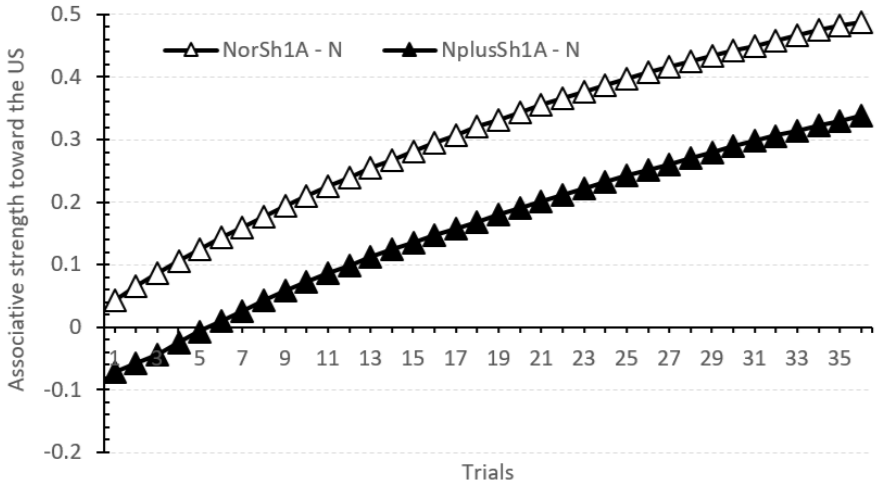


Fig. 3.18 Simulation weights of CS N to the US in the beginning of the third phase of the Baker et. al. 2003 learned irrelevance experiment 1A.

Pre-exposure effects

The following set of experiments all involve pre-exposure, namely the presentation of one or multiple CSs without reinforcement. The relevance of these effects is that they require processes of learning and attention, which decreases the rate with which the exposed CSs are subsequently associated with a reinforcer (latent inhibition, Hall-Pearce negative transfer), or require processing of the differences between non-reinforcing stimuli (perceptual learning). As such, they offer an opportunity to demonstrate the model's presumed mechanisms operating in exposure; namely a decrease in associability (α_r), the formation of context-CS and CS-CS associations through neutral learning, and the effect of these two processes on subsequent reinforced learning.

Latent inhibition and its context specificity

An arbiter of a model's ability to fully explain the effect of latent inhibition, i.e. the attenuation of acquisition due to CS pre-exposure, is whether the model can also produce its context specificity. Models such as PH (Pearce and Hall, 1980) proved very suited to account for the deceleration in latent inhibition, yet fail to explain the observed effect that latent inhibition is weaker when the subsequent conditioning takes place in a different context as compared to pre-exposure. In the DE model, changes in the reinforcer associability α_r account for a deceleration learning rate due to exposure decreasing the magnitude of the time-averaged window of reinforced prediction errors for reinforcers. Just as importantly, by incorporating a predictor error-term in its learning equation, the model postulates that the context specificity of the effect arises from the CS becoming more unexpected when one of its predictors (the context from the pre-exposure phase) is no longer present.

In Experiment 3 of (Channell and Hall, 1983) two groups of rats were given pre-exposure training to a stimulus in a distinctive context and then in a subsequent phase received appetitive conditioning trials to this stimulus. For half of the subjects (Group Exposed Same) conditioning occurred in the same context that was used a pre-exposure whereas for the remaining animals conditioning training was given in a different context (Group Exposed Different). Two further groups of animals (Group Control Same and Group Control Different) received identical conditioning training but did not received pre-exposure to the stimulus. Table 3.9, shows the complete design.

The results of this experiment are displayed on the left panel of Figure 3.19. The conditioning rate was retarded displaying a sigmoidal acquisition shape when the stimulus was pre-exposed in comparison to non-pre-exposed animals (groups Exposed Same and Exposed Different). However, this retardation effect was attenuated when conditioning occurred in a context other than that used during pre-exposure. Thus, Group Exposed Same displayed significantly more latent inhibition, slower conditioning, than Group Exposed Different, which in return showed slight attenuated learning compared to the control groups.

Group	Phase 1	Phase 2
Exposed Same	$A-$*(ctx 1)*	$A+$*(ctx 1)*
Exposed Different	$A-$*(ctx 1)*	$A+$*(ctx 2)*
Control Same	*(ctx 1)*	$A+$*(ctx 1)*
Control Different	*(ctx 1)*	$A+$*(ctx 2)*

Table 3.9 Design of Experiment 3 (Channell & Hall, 1983).

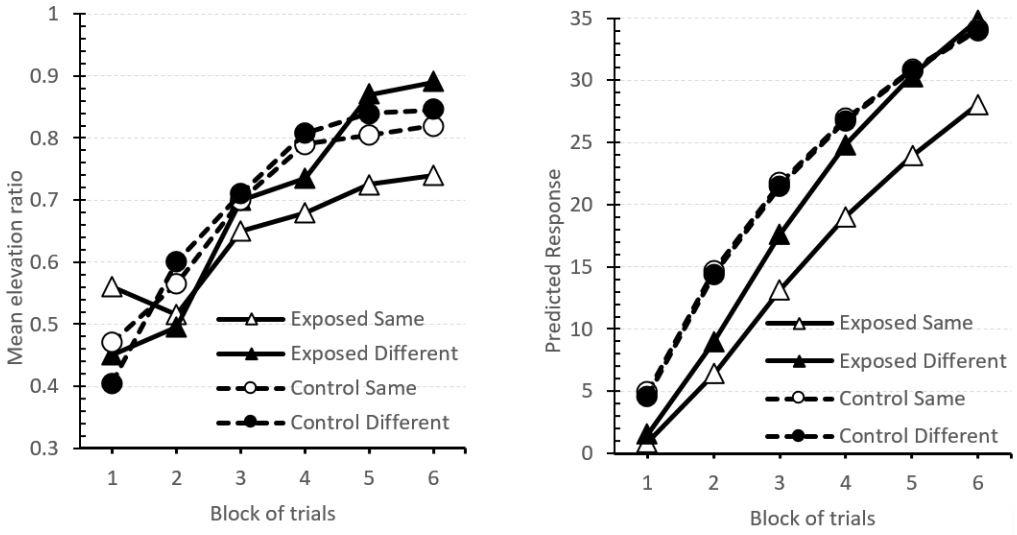


Fig. 3.19 Empirical (original measurement units) and simulated results during the conditioning phase of Experiment 3, Channell & Hall, (1983). The left panel is an adaptation of the paper data showing acquisition of a CR in groups Exposed Same, Exposed different, Control Same and Control different. The right panel displays the corresponding simulated results. Average correlation $R = 0.97$.

The simulated duration of the CSs, US and ITI was 5, 1 and 50 time-units, respectively. The CS salience used was 0.1. Remaining parameters were set as per Table 3.1. Pre-exposure consisted of 100 trials in groups Exposed Same and Exposed Different. Groups Control Same and Control Different received no training in this phase. The second, acquisition phase consisted of 66 trials in all four groups, with the Different groups being programmed to use a different context in this phase. The results of this simulation (Figure 3.19, right panel) closely matched the empirical pattern. Conditioning in Group Exposed Same was delayed in comparison to the other groups. In particular, the effect of pre-exposing the stimulus, latent inhibition, although initially evident was considerable reduced with training when the context was changed from during conditioning (Group Exposed Different).

The DE model replicates the empirical pattern of responses through the combination of its revaluation alpha combined with the second error-term. During pre-exposure, repeated presentations of the exposed stimulus in isolation results in a loss of its associability to the

reinforcer, α_r. As no outcome is expected, there is no uncertainty which, following the model, would be critical in sustaining the associability of a stimulus. Consequently, during subsequent acquisition trials, the stimulus associability to the reinforcer is higher in the groups Control Same and Control Different, because it has not undergone pre-exposure, and therefore keeps its initial associability, than in the pre-exposure groups for which α_r has decayed. In other words, pre-exposure reduces the associability of cues to future reinforcers as they have a history of not being causative of/correlated with reinforcement. Hence, the control groups display faster acquisition partly as a contribution of relatively higher associability towards the US. The model also postulates that the error in predicting the outcome is reduced when the CS is predicted by other stimuli. Thus, in the groups in which pre-exposure and conditioning occur in the same context, the associations between the context and the stimulus and between the CS elements (unitization) formed– during pre-exposure reduce the speed of $CS \rightarrow US$ learning, a decrease which is proportional to the loss of novelty of the CS. As the CS is predicted more strongly in the second phase in Group Exposed Same than in the Group Exposed Different context in which conditioning occurs in a novel context, the rate of learning is further reduced in the former in comparison to the latter, thus resulting in a smaller latent inhibition effect when compared to the corresponding non-pre-exposed control group. In fact, the degree of latent inhibition is postulated to be proportional the net effect of unitization, prediction by the context, and loss of selective attention due to lack of reinforcement.

As these effects can be disassociated, the model can further predict that latent inhibition will be attenuated (yet not completely abolished) should the pre-exposure treatment be followed by pre-exposure to the context alone as this would extinguish the context to CS link and by extension pre-exposure of the context alone prior to the CS presentations. This would be slightly attenuated by such varied pre-exposure raising the uncertainty of neutral cues occurring, and thereby increase the α_n associability of both the CS and context. This

would presumably increase unitization and context to CS learning in the acquisition phase, thereby offsetting the aforesaid attenuation of latent inhibition. An additional prediction is that pre-exposing the CS in multiple contexts should result in stronger latent inhibition, as this would reduce the α_r of the CS, while retaining the strength of the context to CS links. The relative contribution of the revaluation alpha of the model is displayed in a simulation with changes in this parameter disabled (Figure 3.20). This still leads to latent inhibition with context specificity, however the effect is substantially reduced, and the sigmoidal shape of acquisition in the exposed groups is lost. However, a small effect is still observed.

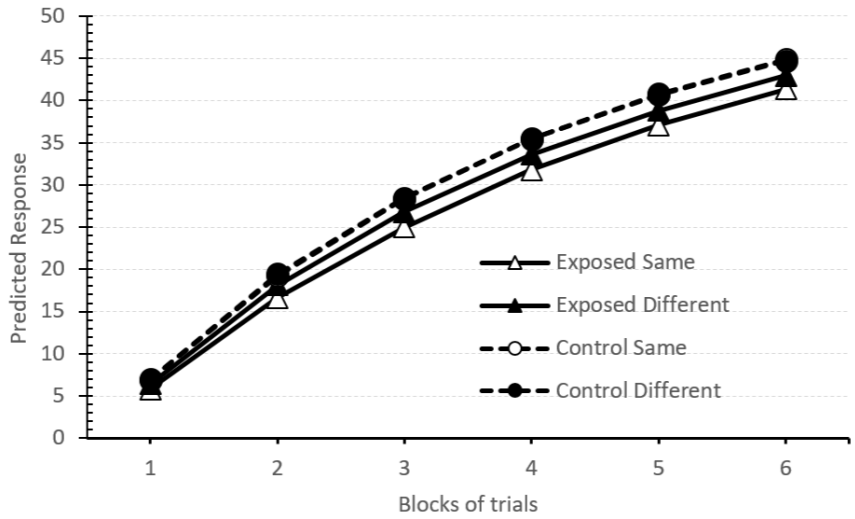


Fig. 3.20 Acquisition in the context specificity of latent inhibition simulation with the α_r associability disabled.

The contribution of the predictor error-term of the model in return displayed in a simulation with associations between neutral stimuli disabled (Figure 3.21). Latent inhibition still occurs, yet the context specificity is lost. Thus, the context specificity arises from unitization and context to CS learning as postulated.

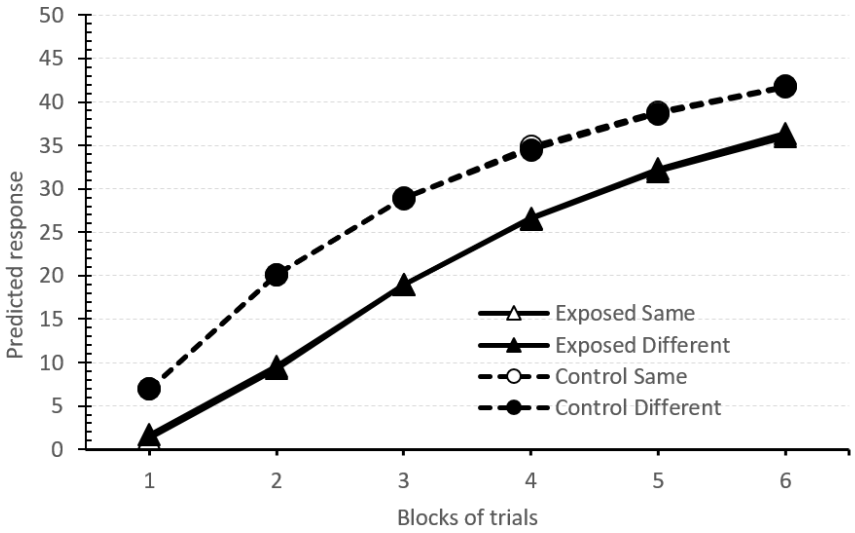


Fig. 3.21 Acquisition in the context specificity of latent inhibition simulation with learning between the context and CS disabled.

As the effects of the revaluation alpha and predictor error-term can be disassociated, the model predicts that latent inhibition will be attenuated (yet not completely abolished) should the pre-exposure treatment be followed by pre-exposure to the context alone as this would extinguish the context to CS link and by extensive pre-exposure of the context alone prior to the CS presentations. However as this decreases the context associability, it is not necessarily evident in the responses during acquisition. Simulating such context pre-exposure after the CS pre-exposure and examining the CS to US weights, it is evident that latent inhibition is indeed attenuated (Figure 3.22).

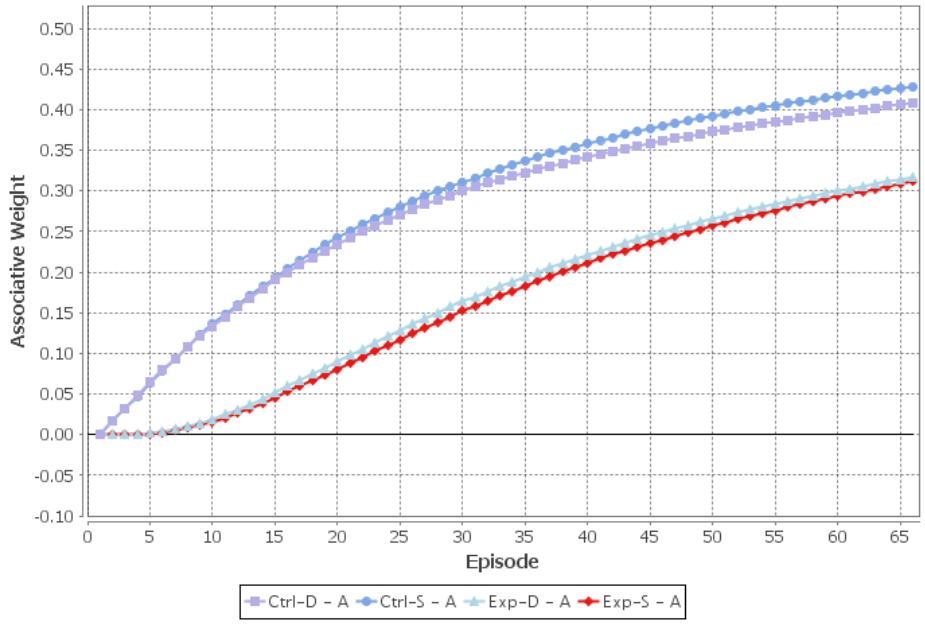


Fig. 3.22 Acquisition in the context specificity of latent inhibition simulation with 100 trials of context pre-exposure after the CS pre-exposure. The result is a decrease in the latent inhibition of stimulus A.

A further prediction is that pre-exposing the CS in multiple contexts should result in stronger latent inhibition, as this would reduce the revaluation alpha of the CS while retaining the strength of the context to CS links. This is borne out by a simulation using pre-exposure in two different contexts (ctx 1 and ctx 3), with acquisition occurring in either ctx 1 for the Same groups, and ctx 2 for the Different groups (Figure 3.23).

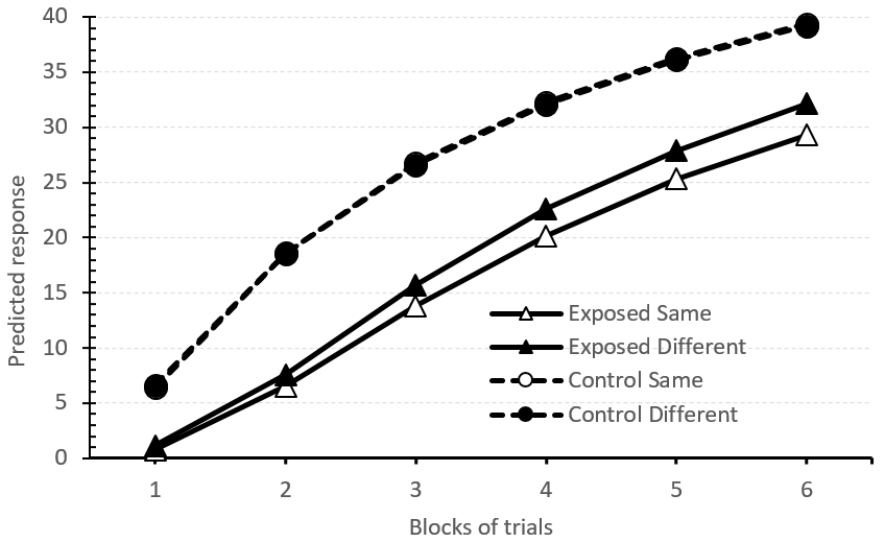


Fig. 3.23 Acquisition in the context specificity of latent inhibition simulation with pre-exposure in two different contexts.

Similarly, a comparative simulation with the Pearce-Hall model failed to produce the context specificity effect, however did produce the general latent inhibition effect (Figure 3.24).

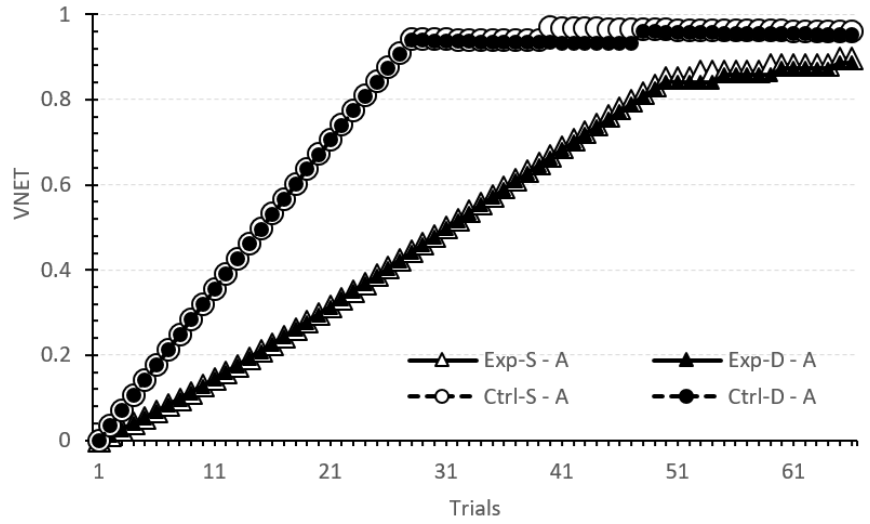


Fig. 3.24 Acquisition in the context specificity of latent inhibition simulation with the PH model simulator (Grikietis, R., Mondragón, E., and Alonso, 2016).

Hall-Pearce effect

To demonstrate how the error-correction and revaluation alpha processes of the DE model can interact to produce emergent effects that seem to run counter to the proposition that acquisition of a CS-US link proceeds monotonically in relation to only the pre-existent associative strength, in this case 'pre-exposure' with a weak US attenuating subsequent acquisition with a stronger US, we simulated the Hall-Pearce effect. Experiment 2 of (Hall and Pearce, 1979) explored the influence of aversive conditioning in rats with a weak shock US upon subsequent learning with a stronger shock CS. In the experiment, two groups of rats received presentations of a tone (Group Tone-shock) or a light (Group Light-shock) followed by a weak shock. The third group, Group Tone alone, received non-reinforced presentations of the tone . In the second phase, all three groups of rats received presentations of the tone followed by a strong shock. The complete design of the experiment is displayed in Table 3.25. The experiment found that reinforcing a CS with a weak US retarded subsequent acquisition towards a more intense version of the same US. This attenuation of learning was however less pronounced than that produced by pre-exposure of a CS. The results of this experiment are displayed in the left panel of Figure 3.26. Group Tone-alone showed a higher suppression ratio than Group Tone-shock, which in return produced a higher suppression ratio than Group Light-shock.

To simulate this experiment, the temporal parameters used were a 5 time-unit CS, 1 time-unit US, and 50 time-unit ITI. Aside from these values, the parameters corresponded to those in Table 3.1. In the first phase, each trial-type was programmed to occur 66 times with a US intensity of 0.1. The second phase likewise consisted of 66 reinforced trials, with a US intensity of 1.0. The simulated results (Figure 3.26, right panel) parallel the experiment. Group Tone-alone showed the largest suppression ratio over the second phase, followed by Group Tone-shock. Group Light-shock produced the lowest suppression ratio,

thus displaying faster learning.

Group	Phase 1	Phase 2
Tone-shock	$66T+$	$66T+$
Light-shock	$66L+$	$66T+$
Tone alone	$66T-$	$66T+$

Fig. 3.25 Hall-Pearce effect design, Hall & Pearce 1979.

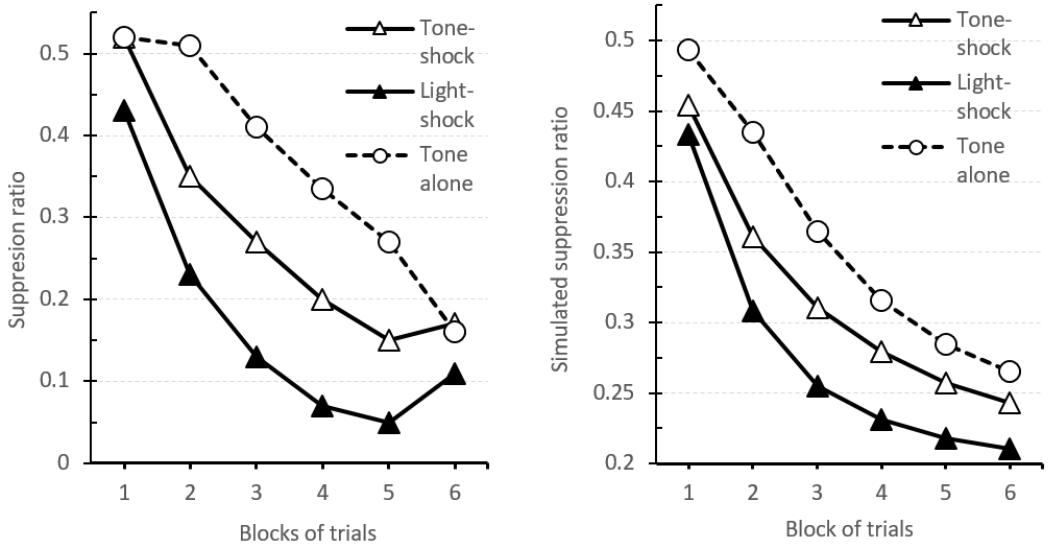


Fig. 3.26 Left: Empirical suppression ratios from phase 2 of Experiment 1, Hall & Pearce 1979 (original measurement units adapted from the paper) showing the acquisition of a CR in groups Tone-shock, Light-shock, Tone-alone, Right: corresponding simulated suppression ratios for phase 2, average correlation $R = 0.97$.

The model accounts for this result due to the Phase 1 treatment leading to a loss of novelty of the tone in both Group Tone-alone and Group Tone-shock, which therefore produces acquisition differences across groups (Figure 3.26, right panel). This loss of novelty is mediated both through the predictor error term of these CSs (when learning about the US) (Figure 3.27), and through a decline in the associability to a reinforcer, α_r (Figure 3.28). However, in the latter group, the CS associability to a reinforcer declines less due to the presence of the shock. In contrast Group Light-shock displays the least attenuation of learning due to

the lack of pre-exposure. Thus, as seen in Figure 3.29, the acquisition curve in the second phase resembles one encountered in latent inhibition. As such, a consequence is that would the experiment be modified to have the second phase treatment occur in a novel context, the Hall-Pearce effect would be significantly attenuated.

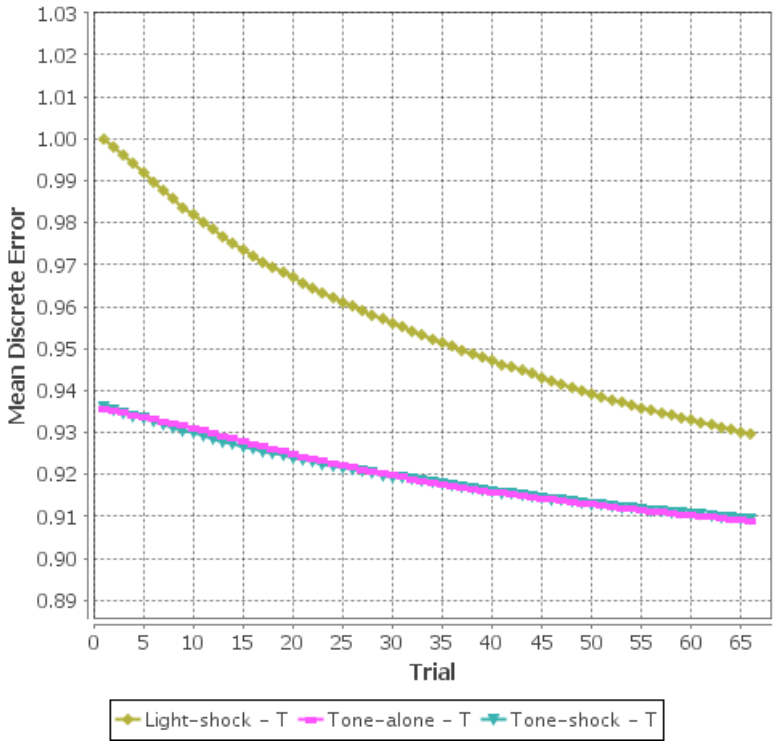


Fig. 3.27 Trial prediction-error for the Tone CS in phase 2 of Hall-Pearce negative transfer.

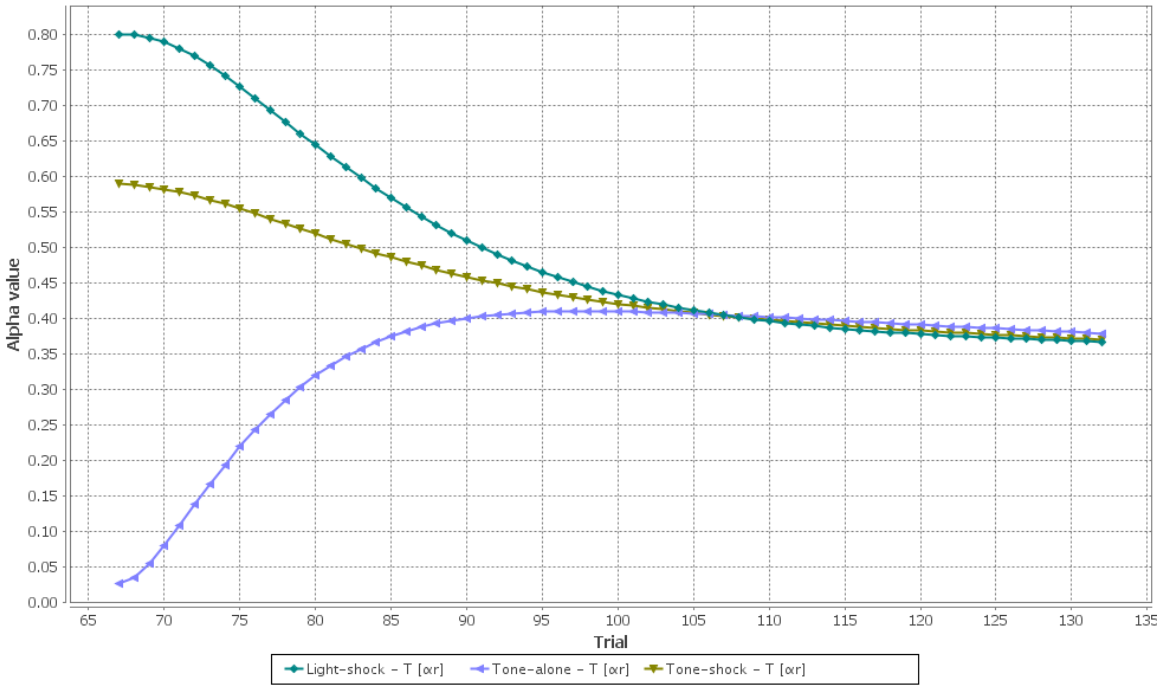


Fig. 3.28 α_r differences for the Tone CS in phase 2 of Hall-Pearce negative transfer.

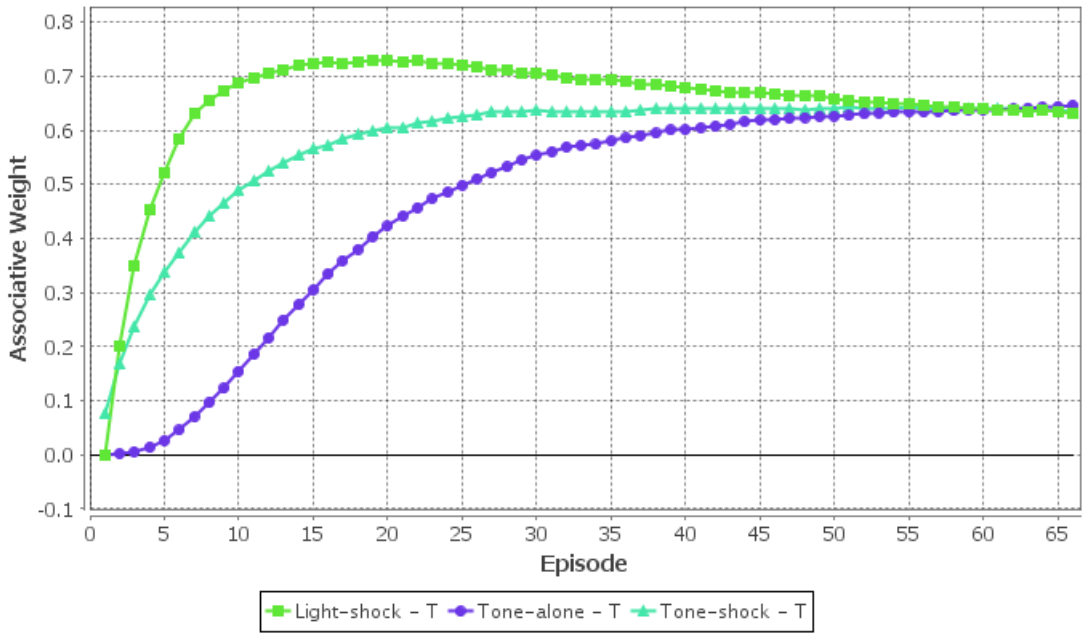


Fig. 3.29 Simulation weights of cue T in the second phase of Hall-Pearce negative transfer, displaying sigmoidal acquisition in the Tone-alone group.

A simulation with such a context change in the second phase is juxtaposed to the standard design in Figure 3.30. As is visible, the effect is indeed attenuated, with both the Group

Tone Alone and Group Light-shock displaying significantly faster learning. Thus, as hypothesized, conducting the second phase of the treatment in a novel phase indeed attenuates the Hall-Pearce effect, as the extant context-CS link from Phase 1, that attenuates the speed of conditioning between the Tone CS and US, is removed.

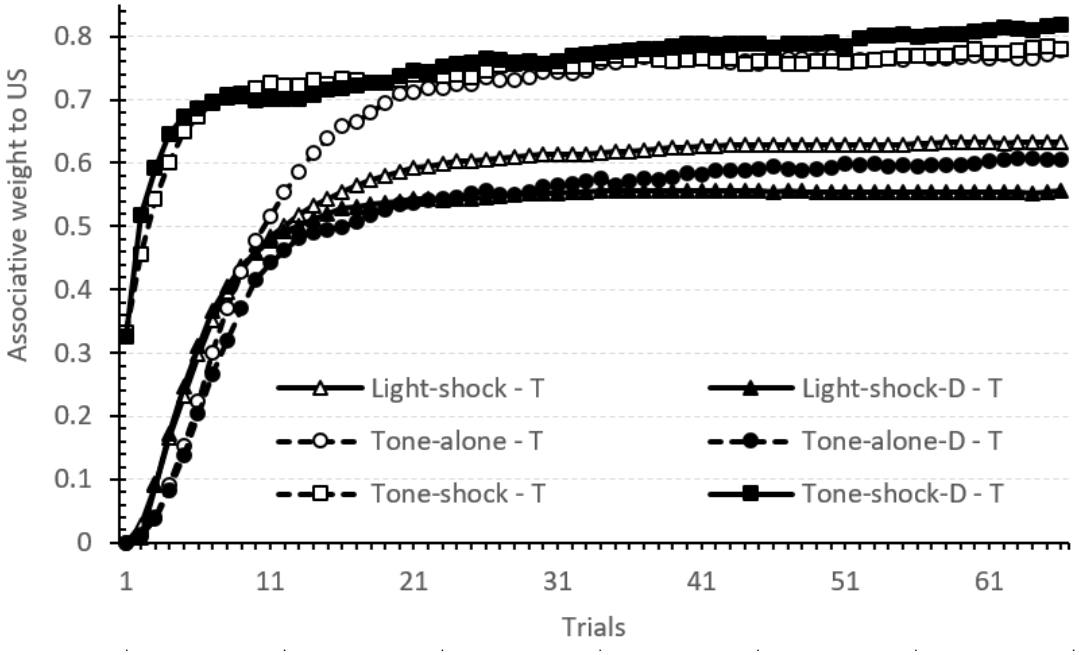


Fig. 3.30 Simulation responding in the second phase of the Hall-Pearce design with either the same context (S) or different context (D).

Perceptual Learning

Perceptual learning is the effect whereby exposure of stimuli facilitates a subsequent discrimination between said stimuli. The phenomenon conveys high relevance because its apparent conflict with latent inhibition: whereas stimulus exposure delays learning, it also facilitates discrimination learning. This facilitation in discrimination has been found to be enhanced when the pre-exposed cues are intermixed (strictly alternated), as compared to being presented in blocks of trials, design intended to control for differences in the amount of exposure and therefore of LI to the cues (e.g., (Hall and Honey, 1989; Mackintosh et al.,

1991; Mondragón and Murphy, 2010; Symonds and Hall, 1995)).

(Blair and Hall, 2003, Experiment 1a), employed a within-subjects design to control further for a differential effect of common stimulus features in assessing the influence of the schedule of exposure. Their experiment, in a flavour aversion preparation, tested PL in generalization test. In Phase 1, rats received non-reinforced exposure to three flavours, compound stimuli AX, BX, and CX. The first half of trials consisted of alternated presentations of AX and BX, followed by a block of CX trials on the second half. This schedule was counterbalanced across animals, so that half of the animals experienced first a block of CX and then the alternated AX and BX trials. On Phase 2, AX trials were followed by a LiCl injection to induce flavour aversion to AX. In the test phase thereafter (Figure X, left panel), consumption of BX was higher than consumption of CX, implying that the aversive learning generalized more to CX than to BX, that is, that the animals discriminated better between the alternated stimuli AX and BX than between AX and the blocked CX. The complete design is displayed in Table 3.31.

Group	Phase 1	Phase 2	Phase 3
PL	$AX-/BX-$ *[intermixed]; CX*$-$	$AX+$	$BX?/CX?$

Fig. 3.31 Perceptual Learning design of Blair & Hall, 2003.

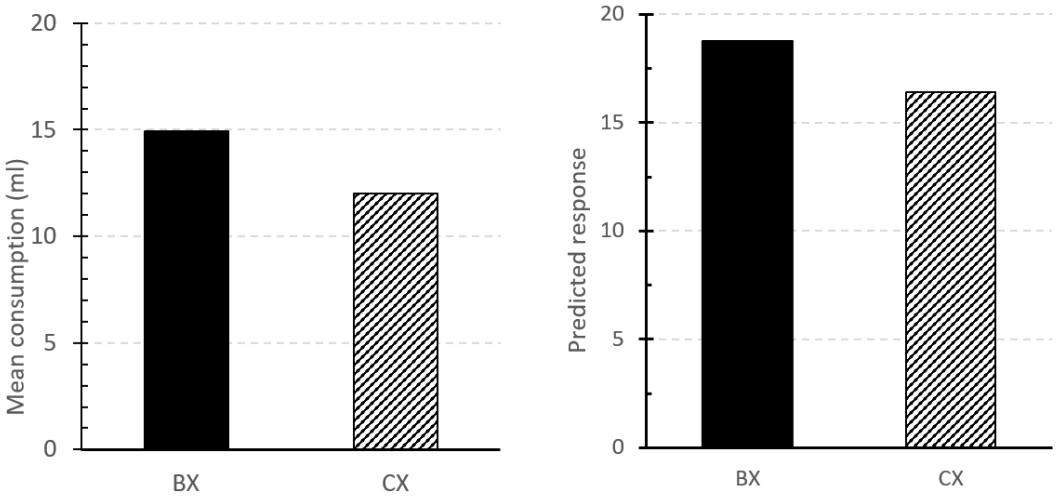


Fig. 3.32 Left: Combined mean consumption (ml) for BX and CX test trials in experiment 1a of Blair and Hall, 2003 (empirical units adapted from paper). Right: Corresponding simulated mean consumption for BX and CX test trials.

A simulation of this experiment was conducted with the following temporal parameters: 5 time-unit compound CSs, 1 time-unit US, and 75 time-unit ITI. The CS and context saliences were respectively 0.5 and 0.1. All other parameters were set per Table 3.1. During Phase 1, each compound was presented ten times in the correct arrangement. Conditioning in Phase 2 consisted of ten trials, and on Phase 3, two generalization trials to BX were programmed. The simulated results (Figure 3.32, right panel) reproduced the empirical data. The simulated response strength as a direct measure of consumption was higher on BX than CX test trials. Consequently, the model was able to predict that intermixed exposure facilitates discrimination when compared to equally exposed blocked stimulus presentation. Per the DE model, intermixed AX, BX pre-exposure results in 1) slightly higher associability in C than B, so that both neutral associations and subsequent reinforcer associations develop at a higher rate; and 2) a stronger $A \rightarrow C$ association than the $A \rightarrow B$ association . Thus, during the AX acquisition trials, mediated conditioning to B will be weaker than to C. Lastly and importantly, intermixed presentations lead to weaker $B \rightarrow A$ than $C \rightarrow A$ links. Thus, during the test phase, the associative chain $B \rightarrow A \rightarrow US$ is weaker in comparison to $C \rightarrow A \rightarrow US$,

contributing towards the disparity between the intermixed and blocked conditions.

However this analysis might lend doubt to whether PL can be obtained by the DE model with reference to a control wherein the control group undergoes no pre-exposure. Thus, I have simulated such a design as well. Phase 1 consisted of 20 trials of BX- in Group BX, 20 trials of CX- in Group CX, 20 trials of AX- in Group AX, 20 intermixed AX- and BX- trials in Group BX-AX, and no training in Group nothing. Phase 2 consisted of 5 trials of AX+ in all groups, and Phase 3 of two trials of BX- in all groups. The design is presented in Table 3.33. The simulation was conducted with equivalent parameters, except a common elements ratio of 0.05 and CS salience 0.1. As is seen in Figure 3.35, responding in Phase 3 for Group BX-AX is significantly lower than in Group nothing, which is not caused by differences in acquisition during AX+ trials as such differences are minute (Figure 3.34).

Group	Phase 1	Phase 2	Phase 3
AX	$AX-$	$AX+$	$BX?$
CX	$CX-$	$AX+$	$BX?$
BX	$BX-$	$AX+$	$BX?$
BX-AX	$AX-/BX-$ *[intermixed];*	$AX+$	$BX?$
Nothing		$AX+$	$BX?$

Fig. 3.33 Second perceptual learning experiment design.

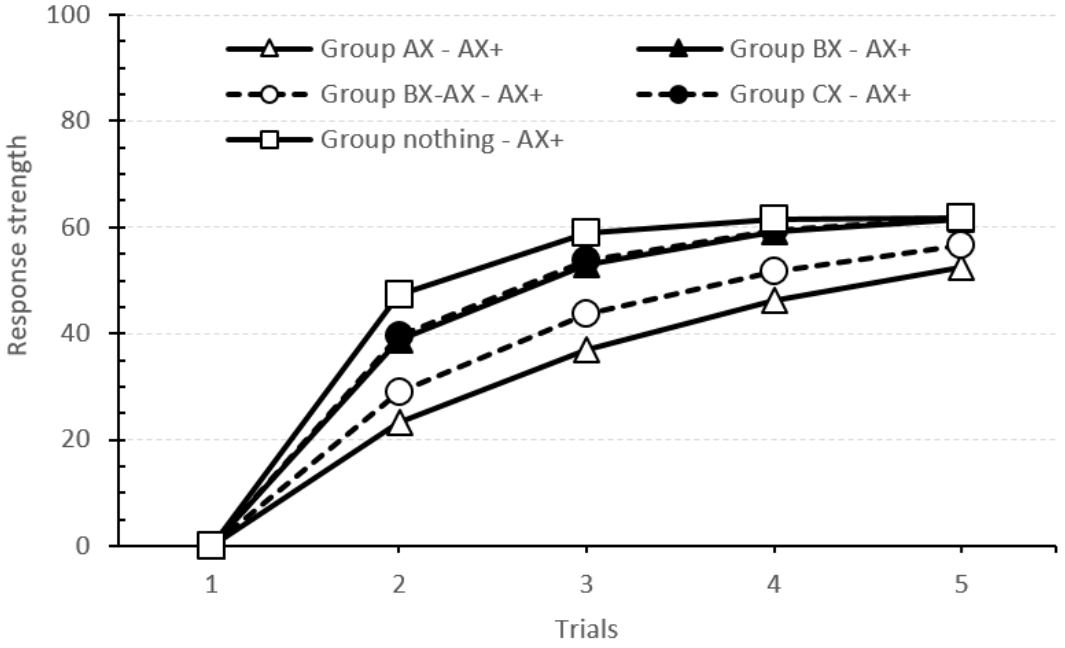


Fig. 3.34 Response strength during Phase 2 of the second PL experiment simulation.

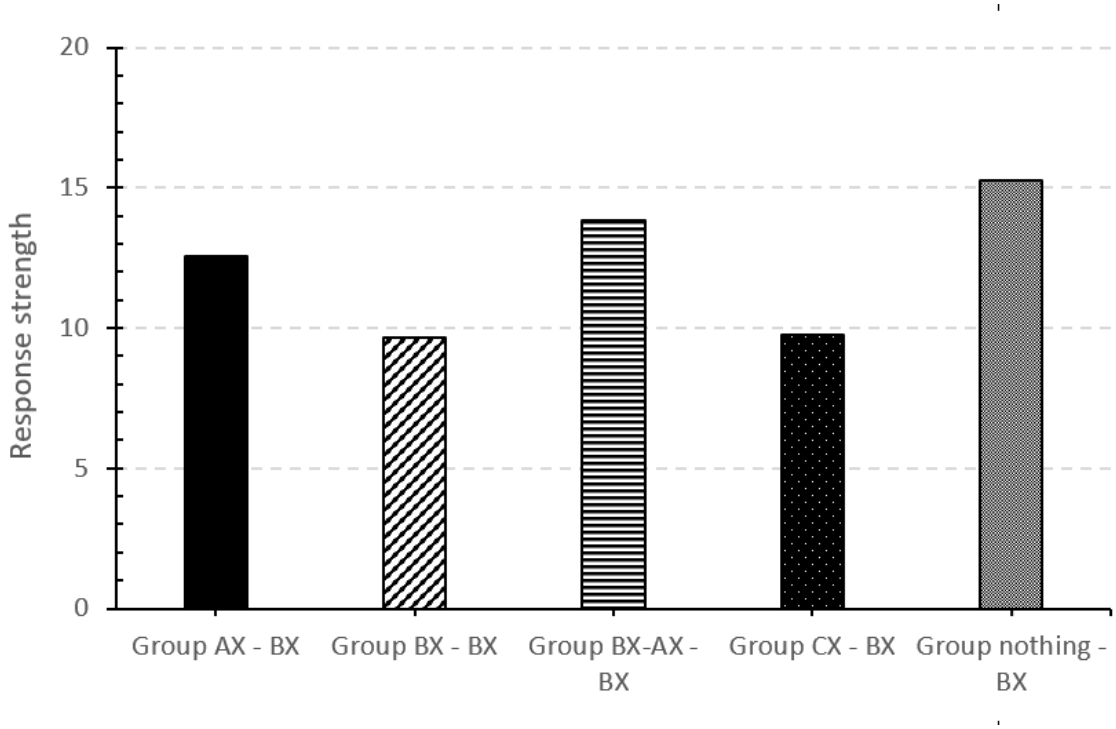


Fig. 3.35 Response strength during Phase 3 of the second PL experiment simulation.

Compound Latent Inhibition

The context specificity of latent inhibition, as we have discussed, lends credence to models of learning that implicate the context in reducing the associability of a pre-exposed CS. In a similar vein, it has been found that pre-exposing a compound of two CSs leads to stronger latent inhibition when a CS member of this compound subsequently undergoes acquisition training. This is of considerable interest, as it implies that the second pre-exposed CS reduces the associability of the CS either directly in the pre-exposure phase, and/or through associative retrieval in the acquisition phase. As such, it supports the case for the DE model's learning rule, which incorporates the unexpectedness of a predictor (in this case the CS) in the rate of learning between a predictor and outcome. That is, we argue the second CS is associatively retrieved during the acquisition phase of the procedure and by proxy acts as a predictor of the retrieving CS, thus reducing its rate of acquisition. Of note is that this effect seems to run counter to the context specificity of latent inhibition detailed earlier, as removing the compound cue could be conceived of as acting as a contextual change. Hence, one might expect that in fact the observed empirical effect should be the opposite. The reason why this is not however the case, might be that the context undergoes more unitization than the CS used in this experiment, the weaker salience of the context as compared to the companion CS, and finally the extinction undergone by the context toward the CS during the ITI in a context specificity latent inhibition design.

Experiment 3 of (Leung et al., 2011), assessed, using aversive conditioning in rats, a novel prediction of the Hall–Rodriguez model (Hall and Rodriguez, 2010) according to which latent inhibition to a target CS pre-exposed in compound with another stimulus should be larger than when pre-exposed in isolation. In the first phase, the two groups of rats were presented with non-reinforced random presentations of A, B, and C. The second phase consisted of non-reinforced presentations of CS compound AB and C. In Phase 3, Group Ele-

ment received reinforced presentations of C (the elemental control), while Group Compound received reinforced presentations of A (the target). The experiment found that pre-exposing A in compound with B led to a more pronounced attenuation of subsequent conditioning, a more robust latent inhibition effect, between A and the outcome when compared to a C that was pre-exposed in isolation. Figure 3.37, left panel, Group Element displayed a higher mean percent of freezing in the test phase than Group Compound, i.e. faster acquisition.

Gr.	Ph. 1	Ph. 2	Ph. 3	Ph. 4
Elem.	$10A-/10B-/10C-$[random]	$10AB-/10C-$[random]	$C+$	$8C?$
Comp.	$10A-/10B-/10C-$[random]	$10AB-/10C-$[random]	$A+$	$8A?$

Fig. 3.36 Leung et. al. experiment 3 compound latent inhibition design.

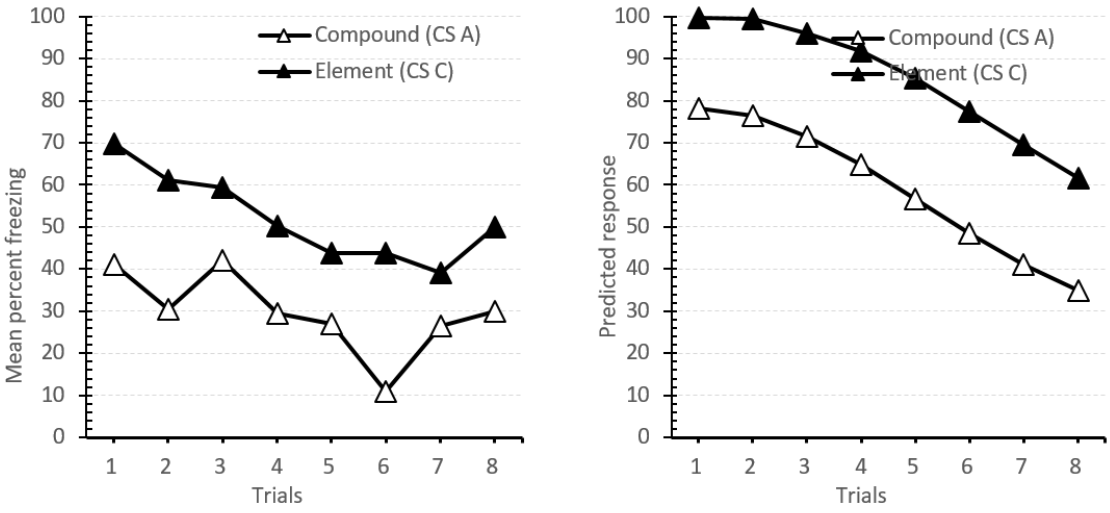


Fig. 3.37 Left: Mean percent freezing in the test phase of experiment 3 Leung et. al. (empirical units adapted from paper). Right: corresponding response strength per trial in the test phase in simulation of experiment 3 of Leung et. al. Average correlation $R = 0.75$.

The temporal parameters of the simulation used a 5 time-unit CS, 1 time-unit US, and 50 time-unit ITI. The US intensity used was 5.0. The CS α_n starting value was 0.8, and the CS and Ctx. saliences both 0.9. The remaining parameters were set to the values in Table 3.1. Both groups of the design were programmed to receive 10 randomly presented

non-reinforced trials of each stimulus (A, B, C) in the first phase. In the second phase, both groups were programmed to receive 10 randomly presented AB- and C- trials. In the third phase, Group Compound received one reinforced presentation of stimulus A, while Group Element received one reinforced C trial. The programming of the subsequent test phase presented respectively 8 nonreinforced A and C trials for groups Compound and Element. The full experimental design is in Table 3.36. The simulation reproduced the general effect observed in the experiment, with similar trends to the empirical responding. Figure 3.37, right panel, Group Compound attains a significantly lower response strength than Group Element in the test trials of the last phase, thereby displaying more latent inhibition. The mechanism operating here, per the DE model, is that the neutral learning between the constituent CSs of the compound AB in the first phase induce the two CSs to become predictors of one-another. This leads to a decreased novelty for the pre-exposed compound, which is reflected in the α_n variable neutral learning associability (Figure 3.38). This further leads to A retrieving B during the second phase in the Group Compound, which in return produces a second-order prediction for A. Since the DE model assumes a predictor error-term in the equation determining the change in the associative weight of a stimulus, this implies that the less novel A in the Group Compound undergoes less acquisition in the second phase than C in the Group Element. That is, learning associations between novel events (a novel predictor and outcome) is prioritized as compared to familiar predictors and novel outcomes.

A further prediction of the model is that should B undergo a series of non-reinforced presentations after being paired with stimulus A, but before the reinforcement of A, this should attenuate the effect observed in this experiment, as it would weaken the causal connection between the two CSs, thereby increasing the novelty of B.

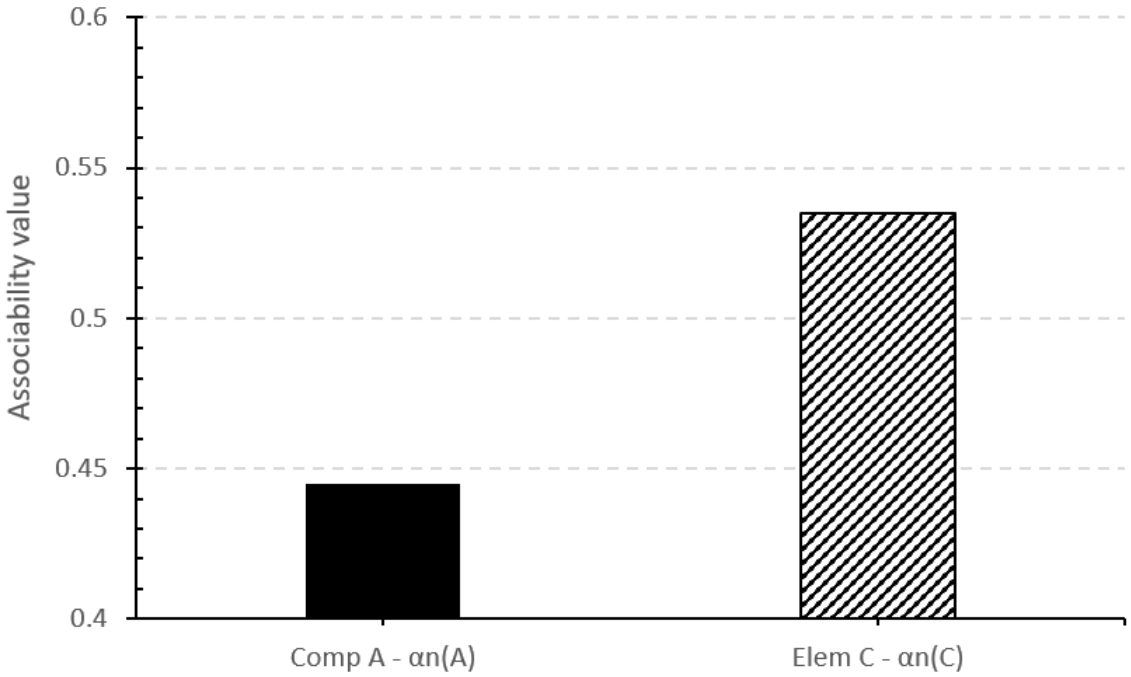


Fig. 3.38 The non-reinforced revaluation alpha, α_n, in the simulated third phase (reinforced trial) of Leung et. al.'s compound latent inhibition experiment. The pre-exposure of two CSs leads to a stronger decline in attentional associability of neutral cues.

In comparison, the PH model predicts no differences between the two groups according to a simulation done with CS alphas 1.0, $\gamma = 0.01$, and $\beta = 1.0$, as seen in Figure 3.39.

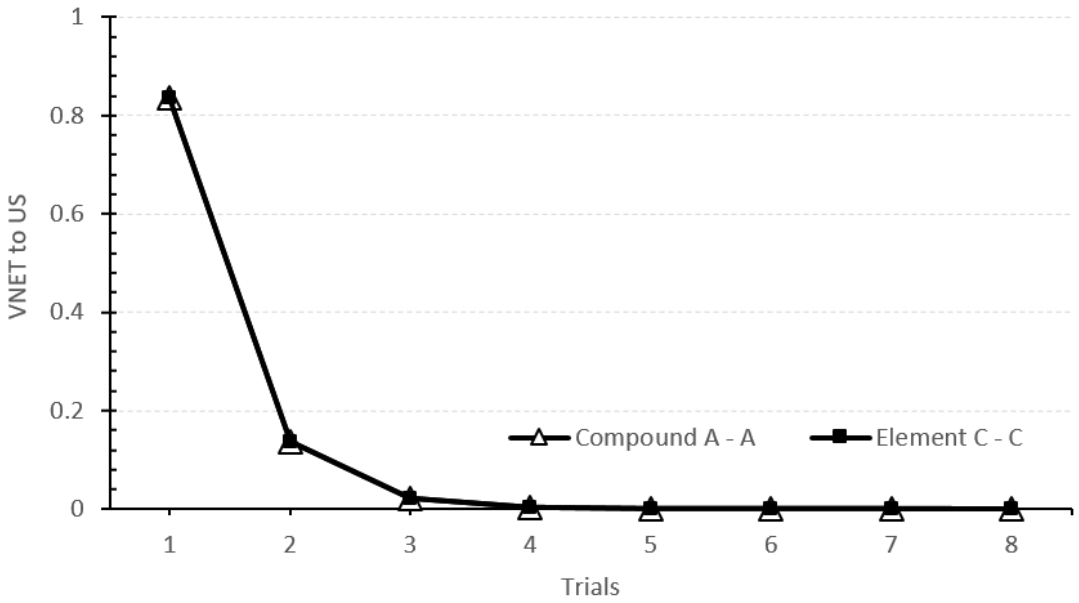


Fig. 3.39 Simulation of compound LI with PH model simulator (Grikietis, R., Mondragón, E., and Alonso, 2016).

Mediated learning

A key feature of the DE model is its ability to account for the way associations change between cues when one or both of which may be associatively retrieved yet physically absent. It accomplishes this through both assuming retrieval of cues by other associatively linked cues and its dynamic asymptote. The latter works on the assumption that the similarity in the level of activation of the predictor and predicted cue dictates the maximal strength of the association that will form between them. Accordingly, a discrepancy in the levels of activity of the stimuli limits their ability to enter into association. This simple idea of adaptability of the asymptote of learning is critical in explaining the apparent contradictory results of mediated related effects, as mentioned in the literature review.

For instance, in retrospective revaluation experimental settings two stimuli, a paired and target CS, undergo reinforced training in compound. After this training is completed, the paired CS is further trained (either in a reinforced or non-reinforced context) independently. As a consequence of this treatment, the associative strength of the paired CS is adjusted, but more importantly, the strength of the target stimulus that does not receive further training is also modified.

Backward blocking and un-overshadowing (often referred to as retrospective revaluation) are exemplary cases. In a backward blocking procedure, following reinforced compound training, the paired cue is subsequently trained with the same reinforcer. As a consequence, the target cue losses some of its initial strength. In an un-overshadowing design, the paired cue is presented in extinction instead, and as a result of this further un-overshadowing treatment, the associative strength of the target cue increases. Formally identical treatments such as sensory preconditioning (SPC) and mediated extinction, in which, the initial compound training is giving in the absence of a reinforcer, have produced results that opposed the

predictions of the theoretical approaches that are able to account for retrospective revaluation experiments. For instance, in SPC subsequent reinforced training of the paired cue results in the target acquiring associative strength, rather than losing it as in the backward blocking procedure. In a mediated extinction procedure, following non-reinforced training of a compound pair-target, the target is conditioned. In a subsequent phase, the paired stimulus receives extinction training. When the target stimulus is next tested a reduction in strength to that attained earlier is observed. This result is opposite to that of the un-overshadowing designs, in which an increase in strength is obtained instead.

Sensory Preconditioning

Sensory preconditioning (Brogden, 1939) is a procedure whereby a compound of two CSs is conditioned in non-reinforced trials. When subsequently one is presented in acquisition trials, the other will elicit a CR when tested. For instance, stimulus *A* is paired with stimulus *B* either serially or as a compound. B is then serially paired with a shock. In subsequent tests, A elicits a CR. This CR is stronger when A & B are treated as a simultaneous compound.

We have tested sensory preconditioning with the following design. A sensory preconditioning and control group were used. Phase 1 consisted of 5 non-reinforced conditioning trials to a 10s stimulus *A* followed by a 10s stimulus *B*, ($A \rightarrow B$) for both groups; In Phase 2, the sensory preconditioning group underwent 10 reinforced trials with stimulus B ($10B+$), while the control group received 10 non-reinforced trials with novel cue N ($10N$); Phase 3 consisted of a single nonreinforced test presentation of stimulus A ($A?$). Table 3.10 shows the design of the simulated experiment. The timing parameters of the simulation involved a 1 time-unit CS, 1 time-unit US, 10 time-unit ISI, and 50 time-unit ITI. The CS α_n starting value was set to 0.8, and the CS saliences were 0.9. The remaining parameters were set according to Table 3.1.

The simulation reproduced SPC, with simulated responding on the test trials being higher in the SPC group than the control (Figure 3.40). The DE model produces this result, due to the combination of its persistent Gaussian activation curves and its neutral learning/associative retrieval mechanisms. In the first phase of the design, the serial compound presentation of A and B forms an excitatory $B \rightarrow A$ link (Figure 3.42) due to the temporal overlap of A and B activity (Figure 3.41). This subsequently leads to the associative retrieval of the representation of cue A in the second phase. This representation interacts with the US to form a slight $A \rightarrow US$ link (Figure 3.43). Of more importance, though, is the A to B link, which during phase 3 leads to retrieval of the US representation through the associative chain $A \rightarrow B \rightarrow US$.

Group	Phase 1	Phase 2	Phase 3
Sensory Preconditioning	$5A \rightarrow B$	$20B \rightarrow +$	$A?$
Control	$5A \rightarrow B$	$20N \rightarrow +$	$A?$

Table 3.10 Sensory preconditioning design.

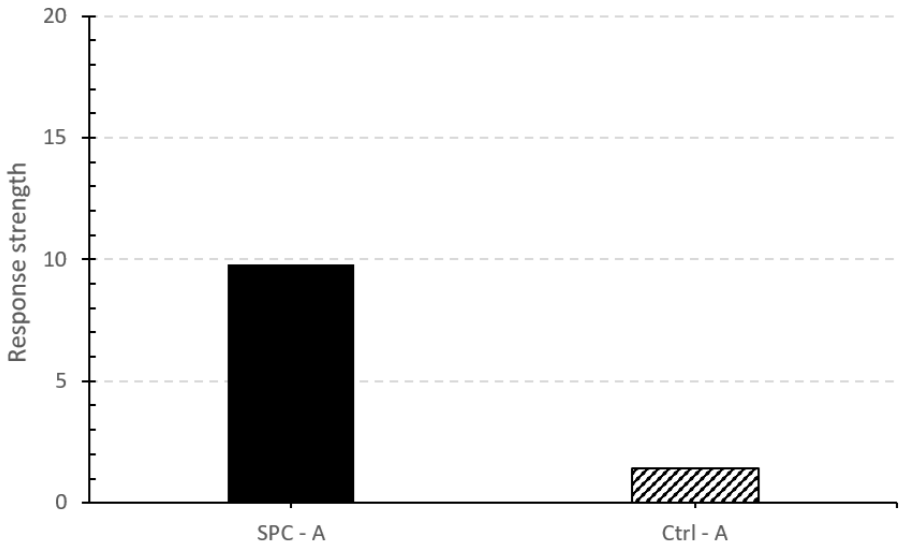


Fig. 3.40 Response strength in the test phase of the simulation of SPC.

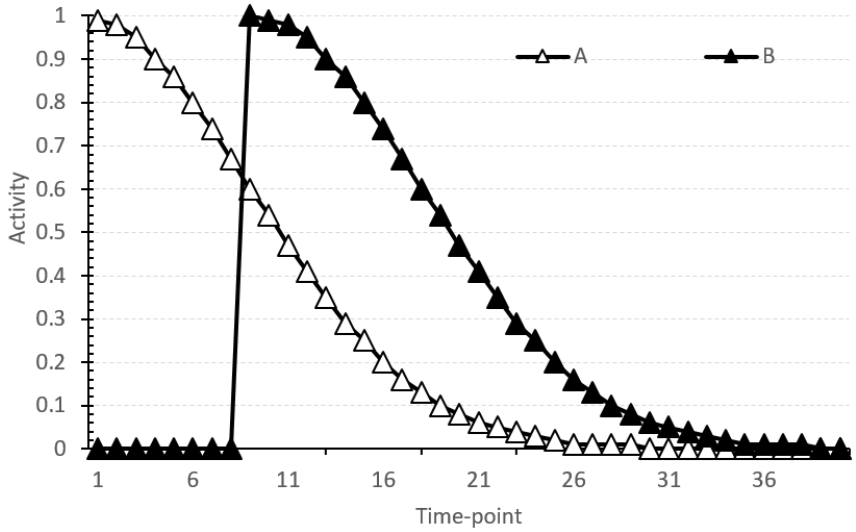


Fig. 3.41 The temporal overlap in activity between cues A and B in phase 1 of the simulation of SPC.

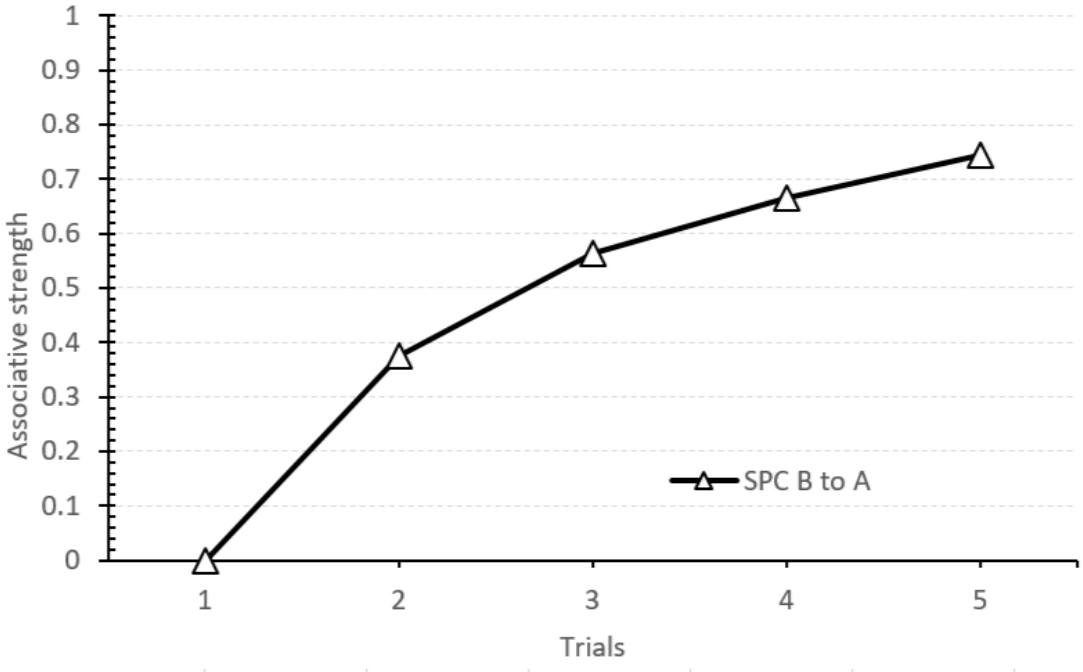


Fig. 3.42 The serial compound presentation of A and B leads to the formation of an excitatory $B \rightarrow A$ link in phase 1 of the SPC simulation.

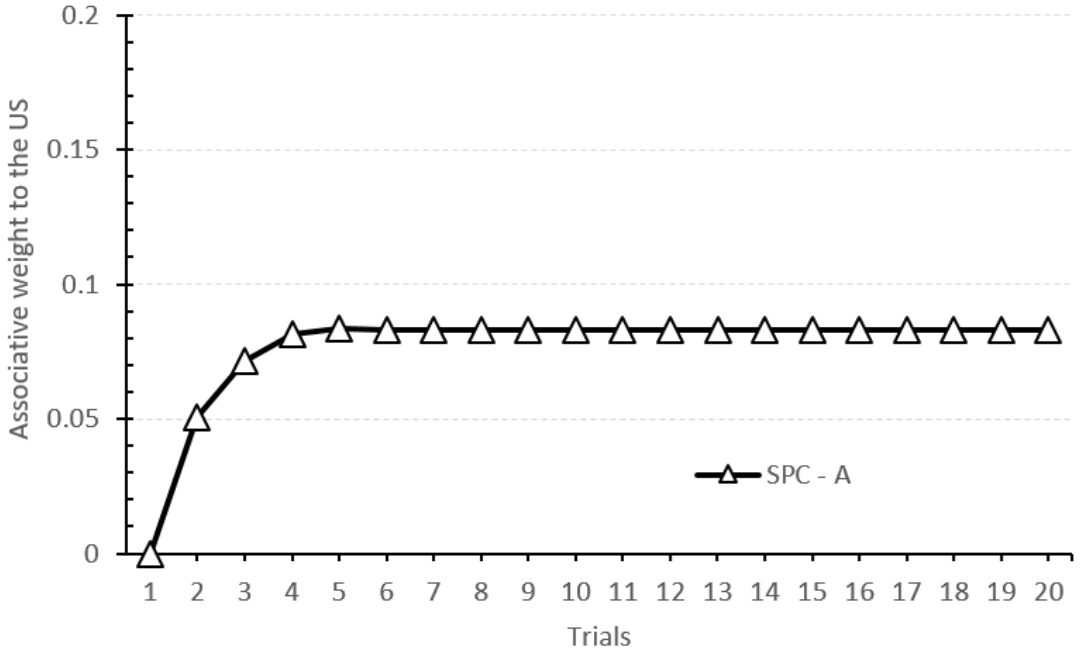


Fig. 3.43 The retrieval of A by B in the second phase of the SPC simulation invokes the formation of an $A \rightarrow +$ link through mediated conditioning.

Backward Blocking

Backward blocking of a cue B is observed in a retrospective revaluation design, when reinforced trials of a CS compound AB are followed by reinforced presentations of A alone, resulting in the absent cue B losing associative strength during reinforced trials of cue A, as assessed by test trials of cue B before and after these reinforced A trials (Wasserman and Berglan, 1998). The DE model predicts this reduction in the associative strength of cue B as occurring due to the within-compound associations between A and B formed during compound training leading to mediated learning when A subsequently retrieves the representation of B on A+ trials. This associatively retrieved representation of B then enters into learning with the present US. As the dynamic asymptote of learning from B to the US, $\lambda_{B \rightarrow US}$, will have a reduced value as compared to trials on which B was present (due to the activities of the two stimuli being more similar on such trials), the mediated learning of B will tend to

reduce its pre-existing link strength to a lower value. Thus the DE model explains backward blocking as arising from the causal connection between B and the US being re-evaluated through an activity-dependent mechanism. Since the prediction of the retrieved cue B for the outcome will be in proportion to its activity, the effect of the number of AB+ trials in the first phase of a backward blocking procedure upon the resultant degree of backward blocking will depend, in the DE model, upon the exact temporal overlap of all cues. For instance, with a 5 time-unit CS and 1 time-unit US, backward blocking is more pronounced with a larger number of trials (Figure 3.44). In contrast, with a 1 time-unit CS and 1 time-unit US the opposite occurs (Figure 3.45). Similarly, the difference between the associative weights of cue A of the 5 and 100 trial groups is significantly larger with a 1 time-unit CS than a 5 time-unit one (see Figure 3.46 and Figure 3.47 for a direct comparison). For both simulations, the CS α_n starting value was 0.8, and the CS and context saliences respectively 0.5 and 0.001. Remaining parameters were set per Table 3.1.

Group	Phase 1	Phase 2
BB5	$5AB+$	$20A+$
BB100	$100AB+$	$20A+$

Table 3.11 Backward blocking design.

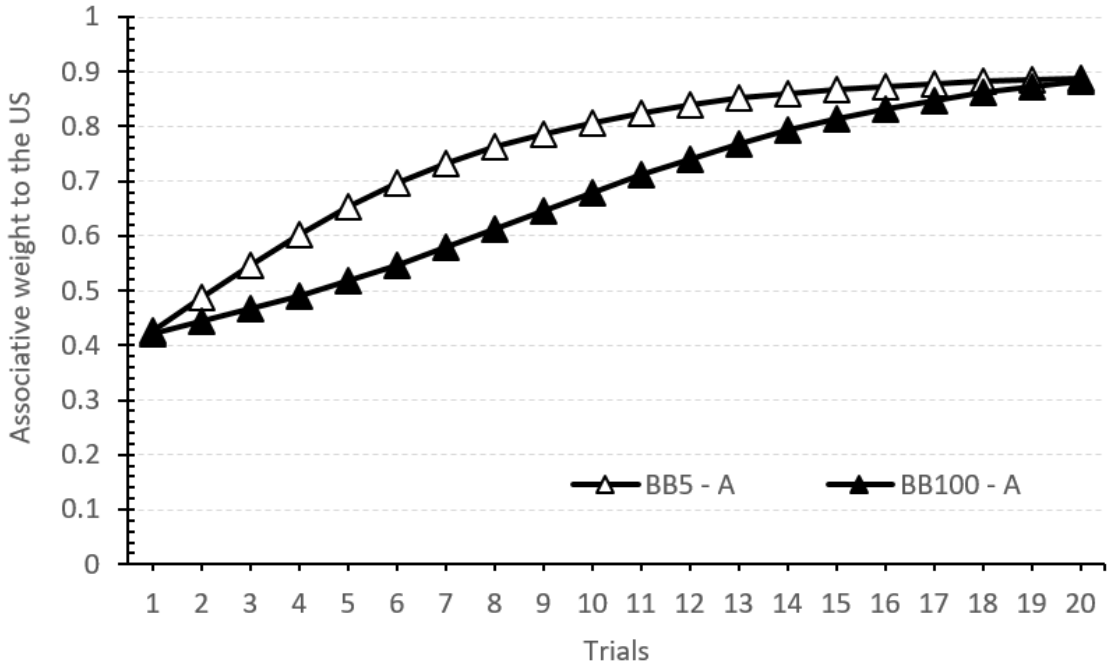


Fig. 3.44 Weights of cue A to the US in the second phase of the simulation of backward blocking with 5 and 100 trials of AB+ training, 20 acquisition trials, 5 time-unit CS, and 1 time-unit US.

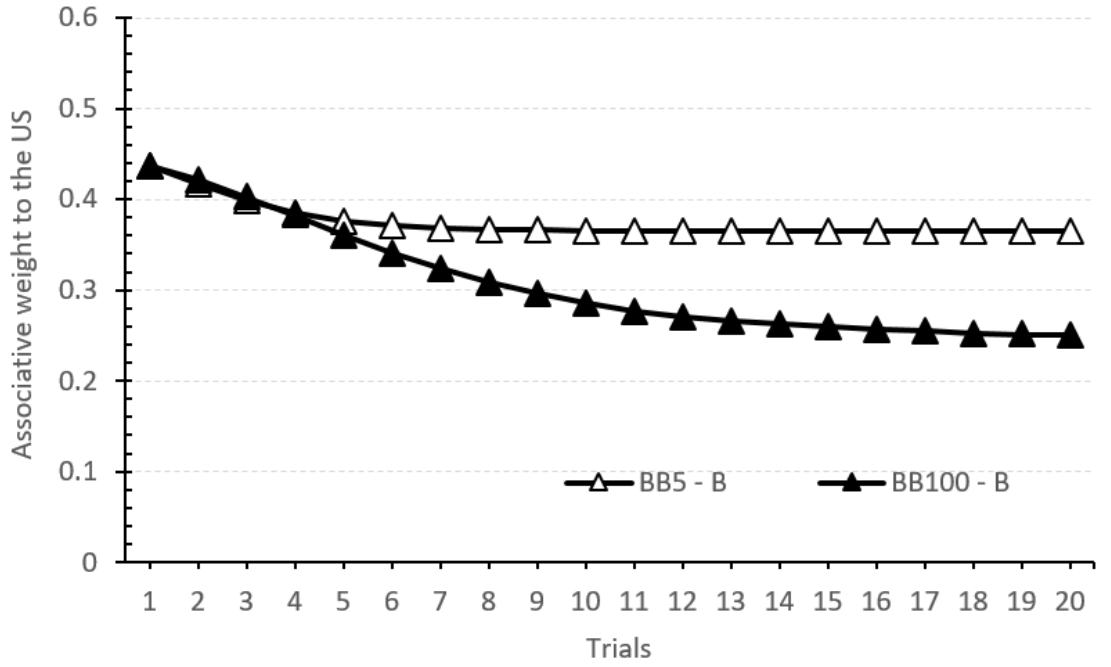


Fig. 3.45 Weights of cue B to the US in the second phase of the simulation of backward blocking with 5 and 100 trials of AB+ training, 20 acquisition trials, 5 time-unit CS, and 1 time-unit US.

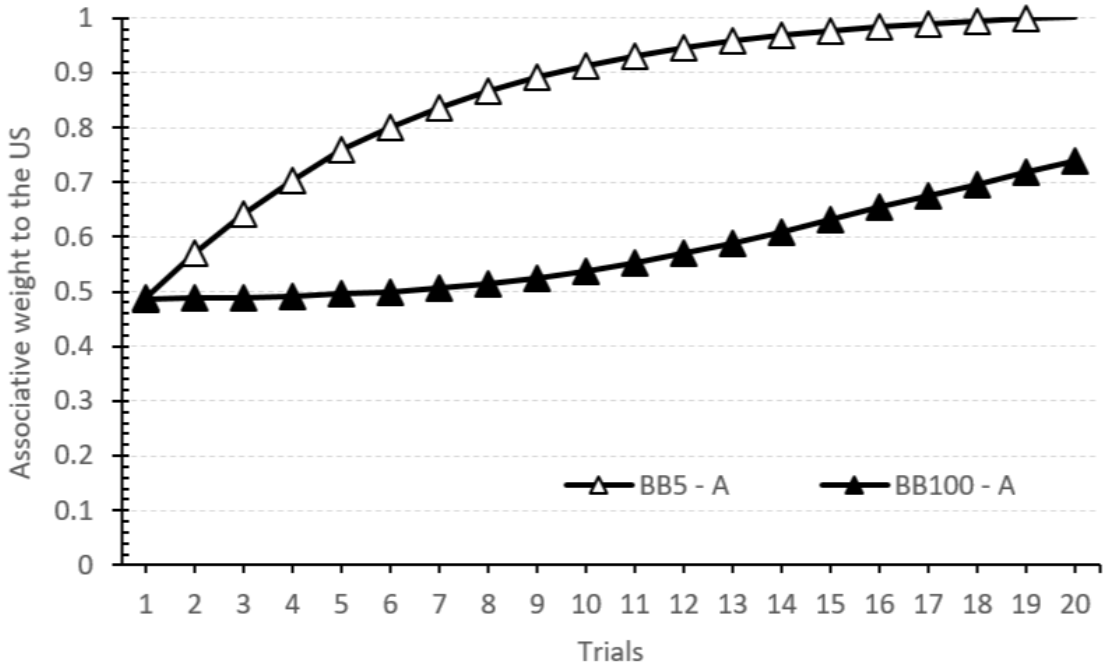


Fig. 3.46 Weights of cue A to the US in the second phase of the simulation of backward blocking with 5 and 100 trials of AB+ training, 20 acquisition trials, 1 time-unit CS, and 1 time-unit US.

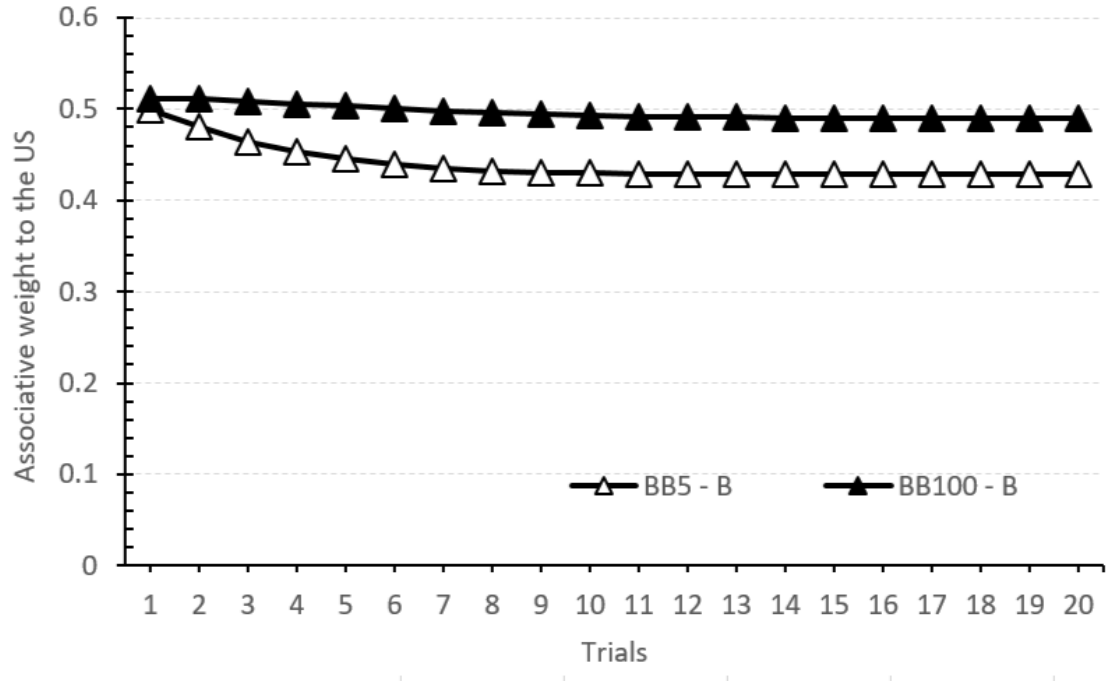


Fig. 3.47 Weights of cue B to the US in the second phase of the simulation of backward blocking with 5 and 100 trials of AB+ training, 20 acquisition trials, 1 time-unit CS, and 1 time-unit US.

Unovershadowing

In contrast to backward blocking, unovershadowing occurs, when reinforced trials of a CS compound AB are followed by non-reinforced presentations of A alone, which results in the associative strength of cue B increasing during the trials on which it is not directly present (San-Galli et al., 2011). The DE model is able to account this effect similarly as in the case of backward blocking. The simulation was conducted with a 5 time-unit CS, 1 time-unit US, and 50 time-unit ITI. The CS α_n starting value was 0.9, while the saliences of CSs and the context were respectively 0.1 and 0.001. Remaining parameters were set per Table 3.1. The model predicts that mutual associations form between cues A and B when they are paired during the first phase of the design. The two CSs similarly form associations toward the outcome. Subsequently in the second phase of training wherein cue A is presented without reinforcement, cue A retrieves a representation of both cue B and the US. Retrieved cue B thence strengthens its associative link toward the retrieved outcome, as their similar activity level entails that the dynamic asymptote of the model supports a high degree of learning between these two cues. This high asymptote in essence means that the two cues have a strong causal connection with one-another. Since, in Phase 1, cue A will tend to form a link toward the outcome relatively more rapidly than toward cue B (due to the outcome being relatively more associable than cue B), the DE model consequently predicts that a larger number of trials in the first phase of this design will tend to facilitate the unovershadowing effect. This is confirmed by a simulation (Figure 3.48), wherein group RR20 underwent 20 phase 1 trials, and in contrast RR100 underwent 100 phase 1 trials. As is visible in the figure, cue B in group RR100 gains significantly more associative strength during the second phase than cue B in group RR20.

Group	Phase 1	Phase 2
RR20	20$AB+$	20A
RR100	100$AB+$	20A

Table 3.12 Unovershadowing design.

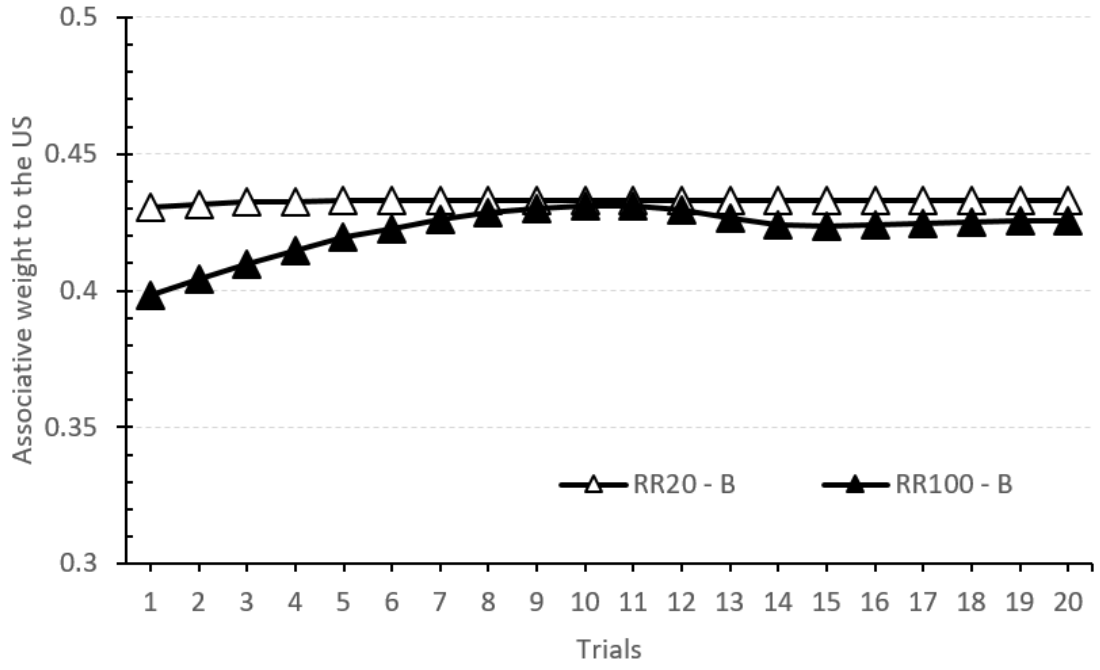


Fig. 3.48 Weights of cue B to the US in the second phase of the simulation of unovershadowing with 20 and 100 trials of AB+ training, 20 acquisition trials, 5 time-unit CS, and 1 time-unit US.

Mediated Conditioning

The phenomenon of mediated conditioning is said to occur when reinforced trials of a CS compound AB are followed by non-reinforced presentations of A alone, which is found to increase the associative strength of the absent cue B, as measured in the discrepancy in responding elicited by cue B before and after the presentations of A with the reinforcer (Ward-Robinson and Hall, 1999). Similarly to retrospective revaluation effects, the model produces these effects due to retrieval of the cue B during the second phase of training by both cue A and the context. That is, the prior pre-exposure of cues A and B elicits the formation of A to B and context to B links which later allow for a retrieval of cue B into

an associatively active state. This retrieved representation of cue B thence interacts with the present reinforcer, forming an excitatory associative link to it. Since the reinforcer and the retrieved cue B have diverging activation levels, the resultant learning asymptotes at a lower level than if cue B were directly present on the these trials. Consequently, the level of mediated conditioning observed will be proportional to the strength of its retrieval, and thus to the quantity of non-reinforced AB trials which precede the acquisition phase of such a design. This is demonstrated by a simulation in which group MC20 underwent 20 phase 1 non-reinforced AB trials, whilst group MC100 underwent 100 phase 1 non-reinforced AB trials. The temporal parameters of this simulation involve a 1 time-unit CS, 1 time-unit US, and 50 time-unit ITI. The CS α_n starting value was 0.8, while the CS and context saliences were 0.9 and 0.2 respectively. The remaining parameters of the simulation are displayed in Table 3.1. As can be seen in Figure 3.49, cue B in group MC100 undergoes slightly more mediated conditioning than cue B in group MC20. This difference is also enhanced by the fact that cue A in group MC100 has a lower α_r value at the beginning of phase 2, thus leaving more opportunity for cue B to form excitatory strength to the outcome than in the other group (Figure 3.50).

Group	Phase 1	Phase 2
MC10	20AB	20$A+$
RR100	100AB	20$A+$

Table 3.13 Mediated conditioning design.

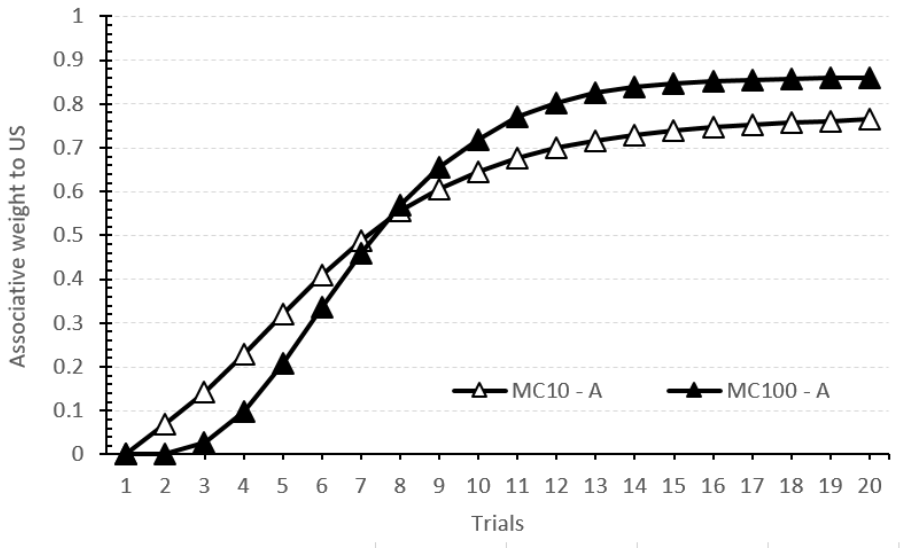


Fig. 3.49 Weights of cue A to the US in the second phase of the simulation of mediated conditioning with 20 and 100 trials of AB training, 20 acquisition trials, 1 time-unit CS, and 1 time-unit US.

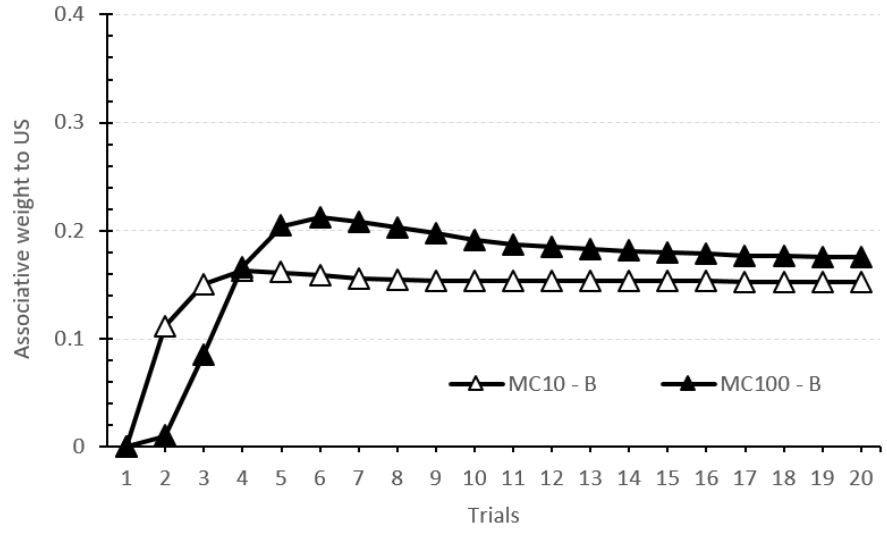


Fig. 3.50 Weights of retrieved cue B to the US in the second phase of the simulation of mediated conditioning with 20 and 100 trials of AB training, 20 acquisition trials, 1 time-unit CS, and 1 time-unit US.

Retrospective Revaluation

In Experiment 4 of (Le Pelley and McLaren, 2001), mediated learning effects were studied in a causal judgement task using human participants, (Table 3.51), constituted by a series of mediated learning conditions and controls. In Condition A2-A2, the first stage consisted of non-reinforced presentations of compound AB. This was followed by reinforced presentations of CS C. In the second stage CS compound AC received non-reinforced presentations. In Condition A2-A1, the first stage consisted of non-reinforced presentations of compound DE followed by non-reinforced F- presentations. The second stage consisted of reinforced DF presentations. Conditions Un-overshadowing and Backward-blocking consisted of respectively reinforced KL and MN presentations. This was followed by respectively non-reinforced presentations of K and reinforced presentations of M. The experiment found no significant mediated learning between an associatively retrieved cue, which had not been reinforced before. This could be due to the cues B and E acquiring context-mediated inhibition during Stage 1 due to the reinforced presentations of cues in that stage, or due to the within-compound associations between the target and competing cues being insufficiently strong. Figure 3.52, top panel, Condition A2-A2 and A2-A1 produced a minute effectiveness rating for the target cues. Condition Un-overshadowing and Backward blocking led to respectively a significantly higher and significantly lower rating for the target cue as compared the controls, Condition RR control and W& B control. This pattern was predicted most effectively by the APECS model (McLaren, 1993) used in the paper.

Condition	Stage 1		Stage 2
A2-A2	$AB?$	$C+$	$AC-$
A2-A1	$DE?$	$F-$	$DF+$
Control	$GH?$	$I+$	$GJ-$
Unover.	$KL+$		$K-$
B. block	$MN+$		$M+$
RR control	$OP+$		$O+/O-$
W&B control	$QR+$		
Fillers	$S-$ $VW-$ $Z\alpha-$	$TU-$ $XY-$	$S\beta+$

Fig. 3.51 Design of experiment 4 of Le Pelley & McLaren, 2001.

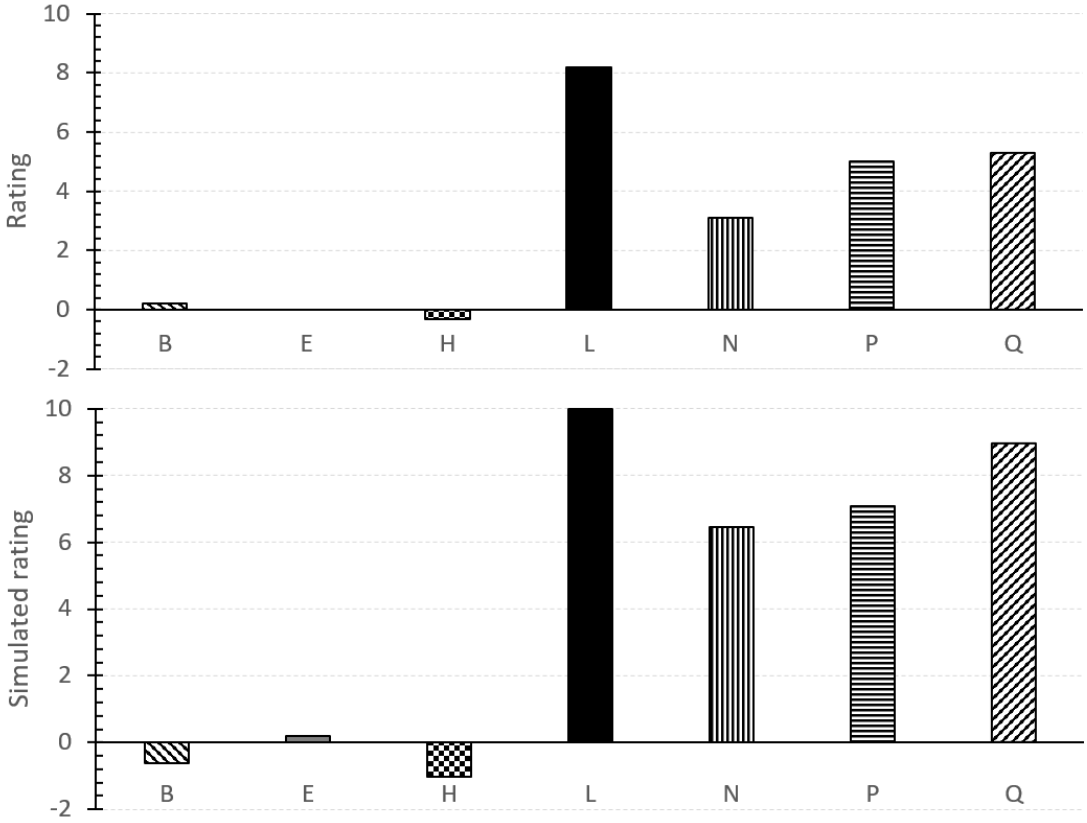


Fig. 3.52 Top: Ratings of target cues after Stage 2 of experiment 4 of Le Pelley and McLaren (empirical units adapted from paper), Bottom: corresponding simulated ratings (derived from the cues' associative weights) after Stage 2 of experiment 4.

The temporal parameters of the simulation used a 5 time-unit CS, 1 time-unit US, and 50 time-unit ITI. The CS α_n starting value was 0.8, and the CS and Ctx. saliences 0.5 and

0.005 respectively. Remaining parameters were set as per Table 3.1. The Stage 1 of the experiment was programmed to consist of 10 trials of each trial-type. The second stage was conducted with 8 trials for each condition. Test trials were programmed with 2 trials of each condition. The full experimental design is in Table 3.51. To calculate simulated effectiveness ratings, associative weights of stimuli to the US in Stage 2 of the experiment were normalized such that the highest weight corresponded to a rating of 10. The simulations matched the empirical data closely. Target cues B, E, and H from Condition A2-A2, A2-A1, and Control all had minute ratings (Figure 3.52, bottom panel). In the case of cue B, the rating deviated in direction from the empirical result. Target cues L and N from Condition Un-overshadowing and Backward blocking were rated respectively higher and lower in simulated effectiveness than target cues P and R from Condition RR control and W& B control.

The DE model predicts that the associative links of cues B, E and H are weak towards the outcome as their Stage 1 non-reinforced presentations occur in the context of reinforced trials of other conditions, thus pushing them towards inhibition/minute excitatory associative strengths. This inhibitory learning is mediated through the slight excitatory strength of the context to the US. Further, the low amount of trials in stage 1 does not allow for significant within-compound links to form between the CS compounds, which impedes Stage 2 mediated learning, by leading to a lower level of retrieval of the target cue. In the second stage of Condition A2-A1, this additionally leads to the dynamic asymptote (measuring the correlation between the two cues) between the weakly retrieved cue E and the present US being so small that CS E effectively undergoes very little excitatory learning.

In the Condition Un-overshadowing, the Stage 1 reinforcement of CS K and L is followed by non-reinforcement of CS K. This leads to excitatory re-evaluation of the L link, as the relative probability of K being a predictor of the outcome diminishes. Specifically, the links

$K \rightarrow L$ and $K \rightarrow US$ from stage 1 of the experiment result in K retrieving L and the + to a similar degree in the second stage. As the $K \rightarrow US$ link extinguishes over these presentations, and as the asymptote in the model supports a high level of learning from L to + due to their similar level of activity (indicating a causal connection between the two events), the net result is a growth of the link $L \rightarrow US$. In the case of Condition Backward blocking, a similar re-evaluation takes place, but in the opposite direction. The link $M \rightarrow N$ that forms in stage 1 of the design leads to M retrieving N in the second stage. As the asymptote of learning in the model between the retrieved N and the present US will be lower than if both cues were present, this results in the extant associative strength of N towards the US decreasing. As the asymptote tracks the causal correlation between the predictor and outcome, this implies that the US and N are causally uncorrelated, hence inducing a weaker link from one to the other. Further, this decrease is dependent on the strength of the associative representation of N as this determines the level of the asymptote of learning between the cued N and the present outcome, as well as the extent to which the retrieved cue N contributes towards predicting the outcome. Hence the model predicts that a larger number of compound trials in the first phase should result in more backward blocking being observed.

Backward Sensory Preconditioning

In SPC pre-training of a target-paired stimulus compound (XA) is followed by conditioning being given to the paired stimulus (A) and by a subsequent test of the strength of the target (X). With a preparation of aversive conditioning in rats, Experiment 2 of (Ward-Robinson and Hall, 1996) aimed to uncover the mechanism underlying SPC, that is, whether sensory preconditioning operated by means of an associative chain $X \rightarrow A \rightarrow US$ or whether a direct link $X \rightarrow US$ was formed through representation-mediated learning. To do so, they employed backward serial presentations of the compound stimuli during the initial training, that is

$A \rightarrow X$, with the intent of preventing the chain association from occurring. Simulating the result of this experiment is an important validator of the capability of the DE model to reproduce conflicting mediated phenomena. The model postulates that a weak excitatory link would be formed between X and A during backward conditioning. Thus, during test, X would still be able to associatively activate the chain $X \rightarrow A \rightarrow US$. Additionally, during conditioning of A, the target X will also be active, and mediated conditioning $A \rightarrow US$ would occur. Mediated conditioning will nonetheless be weak given that the discrepancy between the level of activation of the stimuli involved will bear a low asymptote of learning. The conjunction of both mechanism will therefore be needed to produce the effect.

In the experiment two groups of rats received random presentations of A followed by X, and B followed by Y. In the second phase, presentations of A were followed by an outcome whereas presentations of B were not. In the third phase, Group Ext. received non-reinforced presentations of each A and B, while Group VI received none. The test phase for both groups consisted of X and Y trials. (see Table 3.53). The experiment found that on Group VI, suppression to X that was paired with A which was reinforced in Phase 2 was greater than to Y, paired with B, which in turn was presented without reinforcement. That is, a within-subjects SPC effect was shown in Group VI. Group Ext constituted a further control for the effect. Due to A and B receiving extinction training following condition, no differences in suppression to X and Y were found, indicating that the effect in Group VI relied on the strength of the association between A and the reinforcer. In other words, the extinction treatment to A abolished the backward SPC effect as observed in Group VI. Experiment 3 tested mediated extinction. During Phase 1, animals received identical training to those in Experiment 2. On Phase 2 however, X and Y (rather than A and B) were both reinforced. Oh Phase 3 A was extinguished and the responding to X and Y tested on Phase 4. The results showed mediated extinction of X, that is, a loss of the associative strength of

X as a consequence of extinguishing its paired stimulus A, when compared to the strength attained by Y whose paired stimulus B did not undergo extinction. Figure 3.54 shows the empirical (left panel) and simulated (right panel) results of this experiment.

The temporal parameters used for these simulations were a 5 time-unit CS, 1 time-unit US, and 50 time-unit ITI, the CS α_n starting value 0.8, and the CS salience 0.8. Remaining parameters are presented in Table 3.1. In the first phase, all groups were programmed to receive 12 random presentations of serial-compounds $A \rightarrow X$ and $B \rightarrow Y$. In the second phase, both groups received 4 presentations each of $A \rightarrow +$ and $B-$, presented in the following order ABBA. In the third phase, Group Ext received 44 random non-reinforced presentations of cues A and B, while Group VI was programmed to receive no training. Finally, both groups were programmed to receive 3 random non-reinforced presentations of cues X and Y in phase 4.

Simulated results closely replicated the empirical pattern of responses. In the test phase, suppression to X in Group VI was greater suppression than to Y, this differences were not observed in Group Ext (Figure 3.54, right panel).

Group	Phase 1	Phase 2	Phase 3	Phase 4
Ext	$12A \rightarrow X$ *[rand.]*	$4A \rightarrow US$	$44A - /44B -$ *[rand.]*	$3X?/3Y?$ *[rand.]*
	$12B \rightarrow Y$	$4B \rightarrow US$		
VI	$12A \rightarrow X$ *[rand.]*	$4A \rightarrow US$		$3X?/3Y?$ *[rand.]*
	$12B \rightarrow Y$	$4B \rightarrow US$		

Fig. 3.53 Backward sensory preconditioning design of Ward-Robinson & Hall, 1998.

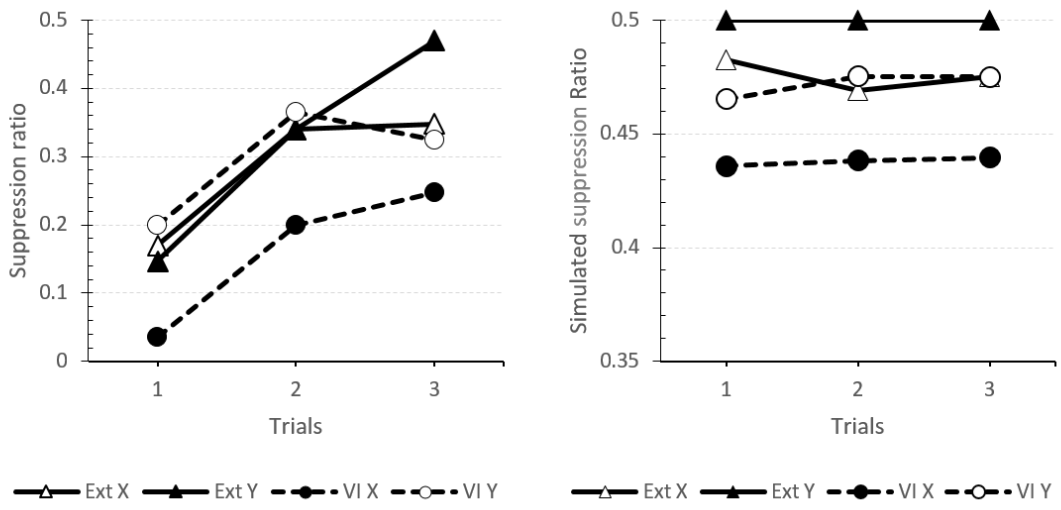


Fig. 3.54 Left: suppression ratios in test phase of experiment 2 of Ward-Robinson & Hall for Group VI and Group Ext (empirical units adapted from paper), Right: corresponding simulated suppression ratio. Average correlation $R = 0.87$.

The DE predicts this effect as arising from the $A-$ extinction trials in group Ext leading to a combination of the mediated extinction of the $X \rightarrow US$ link, representing lack of correlation between these two cues, and of the weakening of associative chain $X \rightarrow A \rightarrow US$ through decay of the $A \rightarrow US$ link. The model predicts that in both groups, the $A \rightarrow US$ link strengthens in phase 1 as CS A is presumed to leave a persistent memory trace even after its offset, which promotes bi-directional excitatory learning between the two CSs. In the subsequent phase, A retrieves X in both groups. Although the less active state of X in both groups implies a lower asymptote of learning towards the present US in the second phase (signifying that B and the US are causally poorly correlated), this nevertheless increases the associative strength of B as its initial strength towards the US in the beginning of the phase is negligible. The non-reinforced trials in Group Ext lead to lower levels of US expectation elicited by cue X due to their effect on extinguishing A. Hence when X is tested, the contribution of the associative chain $X \rightarrow A \rightarrow US$ is diminished. However, predicting this effect does not rely solely on this associative chain. The non-reinforced A presentations diminish the associative strength of X towards the US, as A retrieves the US

more strongly than X. Hence the learning between the weakly retrieved X and strongly retrieved US will tend to be inhibitory due to the dynamic asymptote of the model. Thus, the causal relation between X and the US will be judged to be poor in this scenario. Further, if the model is accurate, it would imply that were the first and second phases of the experiment respectively longer and shorter, it could in fact reverse the effect such that mediated conditioning would take place in group Ext between X and the US. This would occur as elongating the first phase would strengthen the $A \rightarrow X$ link to a level rivalling the more speedily acquired $A \rightarrow US$ association, while shortening the second phase would achieve a similar result by weakening the $A \rightarrow US$ association. This would subsequently lead to X and the US being retrieved to an equally active state on the $A-$ trials, in which case the dynamic asymptote of the model would predict a strong excitatory association from X to the US.

Non-linear Discriminations

Configural, non-linear discriminations arise when reinforced and non-reinforced CS compounds are presented during a conditioning protocol, such that a linear summation of the associative strengths of the constituent CSs is insufficient to accurately predict when reinforcement will or will not occur. These discriminations therefore require the use of additional representational, learning, and attentional processes to introduce non-linearity and thereby solve the discrimination. Despite the elemental trait of the DE model, it is able to solve complex non-linear discriminations primarily through its assumption of elements being able to be activated by multiple stimuli. This effect is accentuated by both the revaluation alpha and predictor error-term of the model. The former leads to elements active on reinforced trials having higher associability. Further, the persistently high error, produced by variable reinforcement, eventually triggers a decay mechanism of the revaluation alpha that disproportionately affects common elements. The latter operates through the network of non-reinforced learning of the model to reduce the learning rate of cues that are predicted well by other cues, thereby allowing cues that are predictive of reinforcement or non-reinforcement to learn faster.

Negative Patterning

(Whitlow and Wagner, 1972, Experiment 1), studied a NP discrimination, in eyeblink conditioning. In the Phase 1, rabbits were presented with reinforced presentations of cues A, B, and C. The cue C was used to demonstrate that though the responding elicited by a compound CS is related to the summation of the associative strengths of its constituent cues, that there are functionally configural components that also contribute towards determining the associative strength of said compound. The next phases consisted of random presentations of reinforced stimuli A and B, as well as the non-reinforced compound AB. Finally, all combinations of the three cues were tested in a random order. The experiment found significant negative

patterning, that is the rabbits learned to withhold responding on AB trials. Further, test showed that this withholding of response was only evident for AB trials (Figures 3.56 and 3.57, left panels).

Group	Days 2-3	Days 4-10	Day 11	Day 12
NP	180x A+/B+/C+	80x A+/B+/2AB-[random]/day	40x A+/B+/2AB-/C+ [random]	6x (A/B/C/AB/AC/BC)?

Fig. 3.55 Negative patterning experimental design of (Whitlow & Wagner, 1972).

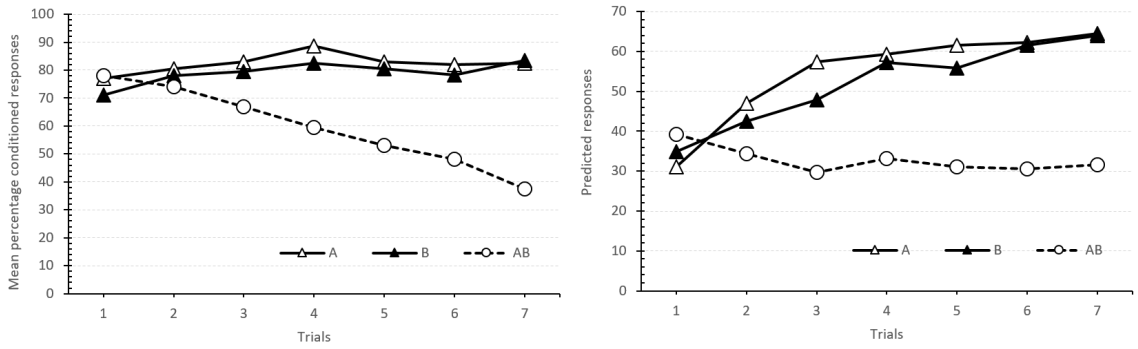


Fig. 3.56 Left: Responses in the discrimination phase of the negative patterning experiment of Whitlow & Wagner, 1972 (empirical units adapted from paper). Right: Equivalent simulated responses.

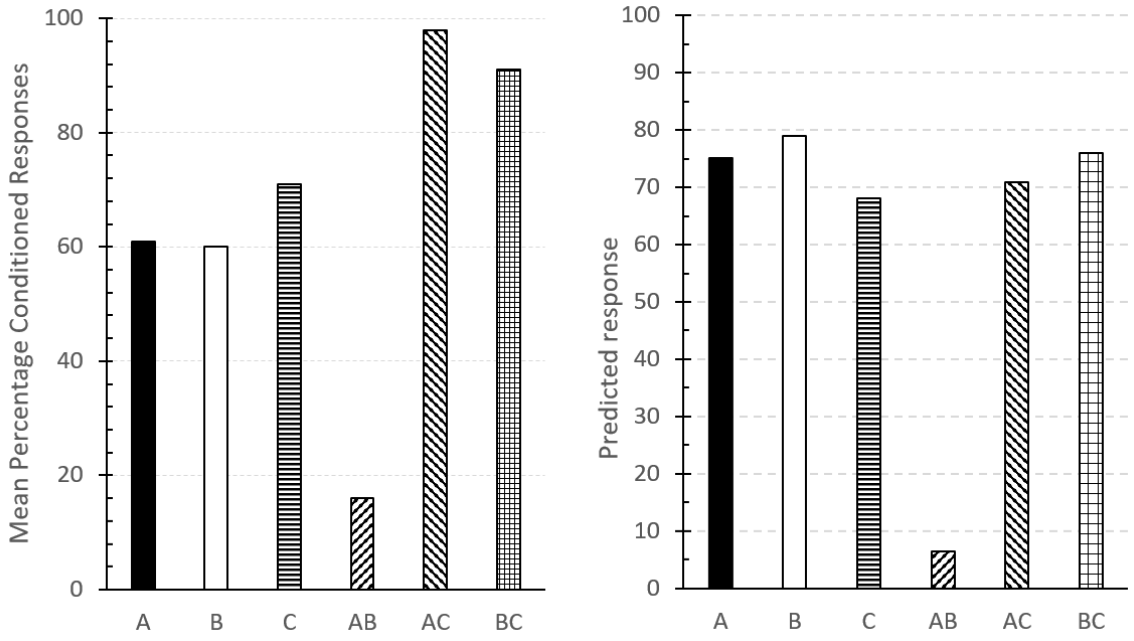


Fig. 3.57 Left: Mean percentage conditioned responses per compound in the test phase of the negative patterning experiment of Whitlow & Wagner, 1972 (empirical units adapted from paper). Right: Corresponding response strength in the test trials of the simulation.

The temporal parameters of the simulation were 1 time-unit CS, 1 time-unit US, and 20 time-unit ITI. The context salience was 0.1, a recency of 0.1 was used, and the common elements ratio was 0.2. The remaining parameters of the experiment are found in Table 3.1. For the first phase, 180 reinforced trials of cues A, B, and C were programmed to be presented in a semi-random manner. Phase 2 consisted of random 560 presentations of reinforced A and B trials and 1120 nonreinforced AB- trials. Phase 3, consisted of random presentations of 40 reinforced presentations of A and B, 20 reinforced presentations of C, and 80 non-reinforced presentations of the compound AB. Finally, the test phase consisted of a single, non-reinforced presentation of trial-types A, B, C, AB, AC, and BC. The simulation closely matches the experimental data. Figure 3.56, right panel, shows the acquisition of the NP discrimination where responding to the compound AB progressively declined across trials. The right panel of Figure 3.57 displays the test results that parallel those found empirically,

i.e., suppression of responding to AB was larger than to all other stimulus or compounds.

According to the DE model during discrimination training, redundant cues (the elements common to A and B as well as the context) become highly excitatory toward the US, while elements unique to A or B become highly inhibitory. Further, on AB trials the common elements between A and B can only be sampled once. Thus, they contribute relatively less to responding than on A or B trials, thereby leading to a lower response being elicited on such trials by these elements than would be expected by linear summation. Two unique mechanisms of the DE model strengthen this effect. Firstly, unique elements of A and B on AB- trials strongly predict the common elements, thereby protecting them from extinction by decreasing their associability. Secondly, due to the unique decay mechanism of the associability to the reinforcer, α_r. Inconsistent outcomes maintain high levels of attention, however, when training under these conditions is prolonged, that is, when sustained consistent high error endures after long training, a mechanism to cease sustained attention is activated. This mechanism becomes active sooner for the common elements than for unique elements, due to these elements being active on more trials and thus accruing evidence faster of the contingency being inherently random. Thus, this further reduces their associability and pushes the unique elements to become more inhibitory. The net effect is that on compound trials, the inhibitory strength of the unique A and B elements is of higher magnitude than the excitatory strength of common elements, thereby leading to a with-holding of response. On A and B trials in return, only half of this inhibition occurs, yet the same quantity of common elements is active. Therefore, responding is higher on these trials.

In contrast, a simulation with the Pearce model did not predict the observed result of the experiment (Figure 3.58).

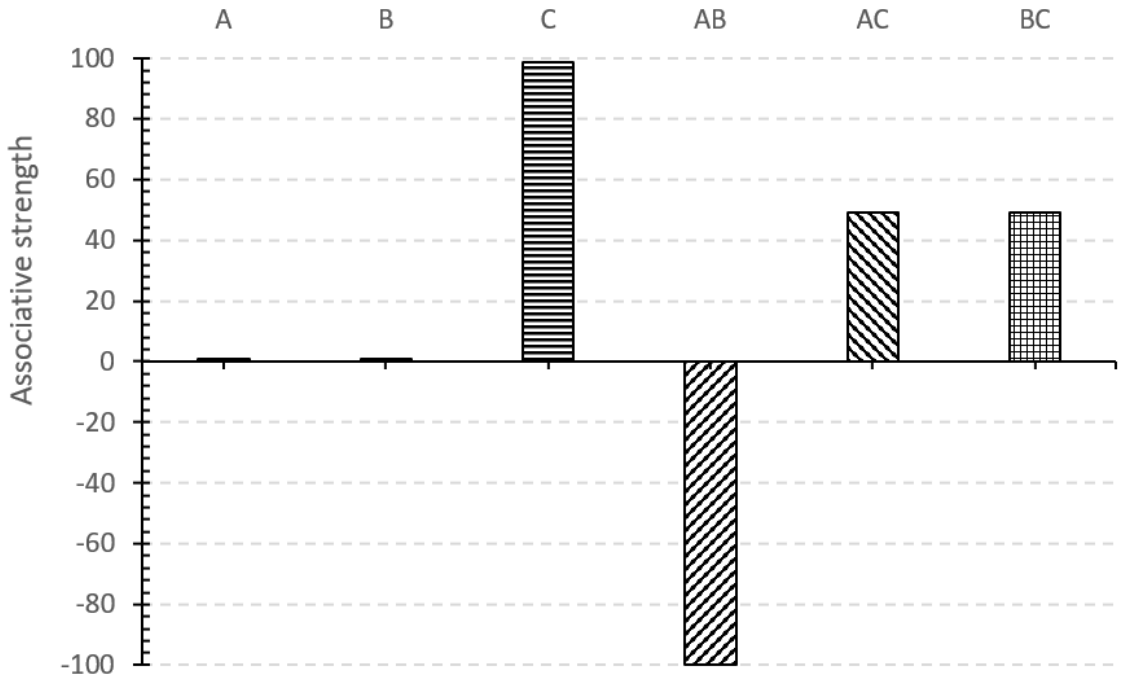


Fig. 3.58 Associative strength in the test trials of the Pearce model (Gheorghescu, A., Mondragón, E. and Alonso, 2017) simulation of Whitlow & Wagner, 1972.

Biconditional Discrimination

In a biconditional discrimination four compounds of two stimuli (AB+/CD+/AD-/BD-) are reinforced in a manner such that each individual cue receives both reinforced and non-reinforced training. This discrimination, from the elemental perspective, is difficult to solve due to each stimulus receiving equal partial reinforcement. It therefore seems to require the existence of configural nodes, which enable the discrimination to be solved through non-linear means. However, the assumption of pairwise common elements in the stimulus representation in the DE model, a mild assumption, circumvents this difficulty; increasing the power of this and other elemental models to predict 'configural' learning without the assumption of a configural representation.

We simulated the biconditional discrimination experiment of (Lober and Lachnit, 2002), which used an aversive skin-conductance conditioning procedure in human participants. For the biconditional treatment, participants received intermixed random presentations of reinforced (AB+/CD+) and non-reinforced (AD-/BD-) compounds. The participants showed a progressive tendency to withhold responding on non-reinforced trials. (Figure 3.59, left panel).

Group	Phase 1
BD	$50AB+/50CD+/50AC-/50BD-[random]$

Table 3.14 Biconditional discrimination design of Lober et. al.

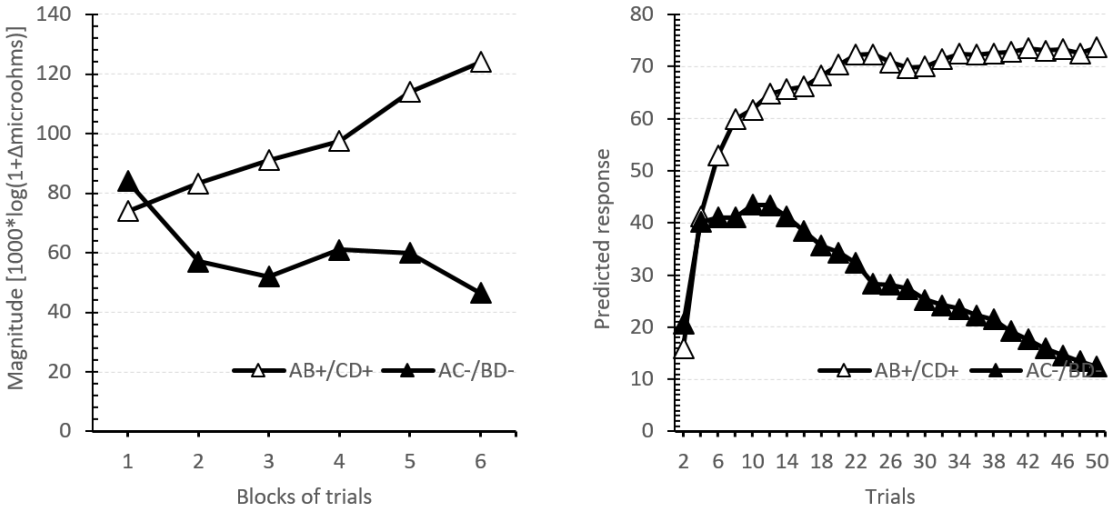


Fig. 3.59 Left: Second-interval responses (SIR) for combined AB+/CD+ trials and AD-/BC- for the biconditional training of Lober and Lachnit, 2002. Right: Corresponding combined simulated response strength in the AB+/CD+ and AD-/BC- trials. Average correlation $R = 0.72$.

The simulation's temporal parameters consisted of a one time-unit CS and US, and a 50 time-unit ITI. The parameter values used are presented in Table 3.1. The first and only phase was programmed to consist of 50 randomly presented trials each of reinforced compounds AB+ and CD+, and of the non-reinforced compounds AC- and BD-. The full design is in Table 3.14. The pattern of learning was reproduced in the simulation. Figure 3.59, right panel, there was an initial increase in responding on both reinforced and non-reinforced trial-types. This was followed by a gradual decrease in responding on non-reinforced trials, with responding on reinforced trials staying stable. This prediction is produced by the model predicting that elements (common and unique) which are predictive of reinforcement gaining excitatory strength, while the reverse occurs in elements predictive of lack of reinforcement. Specifically, through the unique elements of all CSs (Figure 3.60), along with the common elements of AB and CD gaining excitatory strength (Figure 3.61). The latter become excitatory due to their increased sampling on AB+ and CD+ trials. Simultaneously, the common elements of AC and BD become conditioned inhibitors (due to their increased sampling on AC- and BD- trials), while the elements common to AD and BC have weights near zero (Figure 3.61)

(due to their sampling being uniform across trials). As the latter are redundantly sampled on reinforced trials, the concomitant summation effect produces the discrimination. This is further facilitated by the common elements unable to aid in learning the discrimination, AD and BC, decreasing their revaluation alpha as compared to other cues. The contribution of the proportion of common elements towards the discrimination is elucidated by a simulating using 10% common elements instead of 40% as in the previous simulation (Figure 3.62). This still produces biconditional discrimination, however the discrimination is learned more slowly. One might argue that these sets of common elements are rather arbitrarily defined, and we believe this to be true to an extent. A more principled approach would, as mentioned in Chapter 2, involve learning commonalities between stimuli derived from some sensory modality through deep connectionist networks (Mondragón et al., 2017).

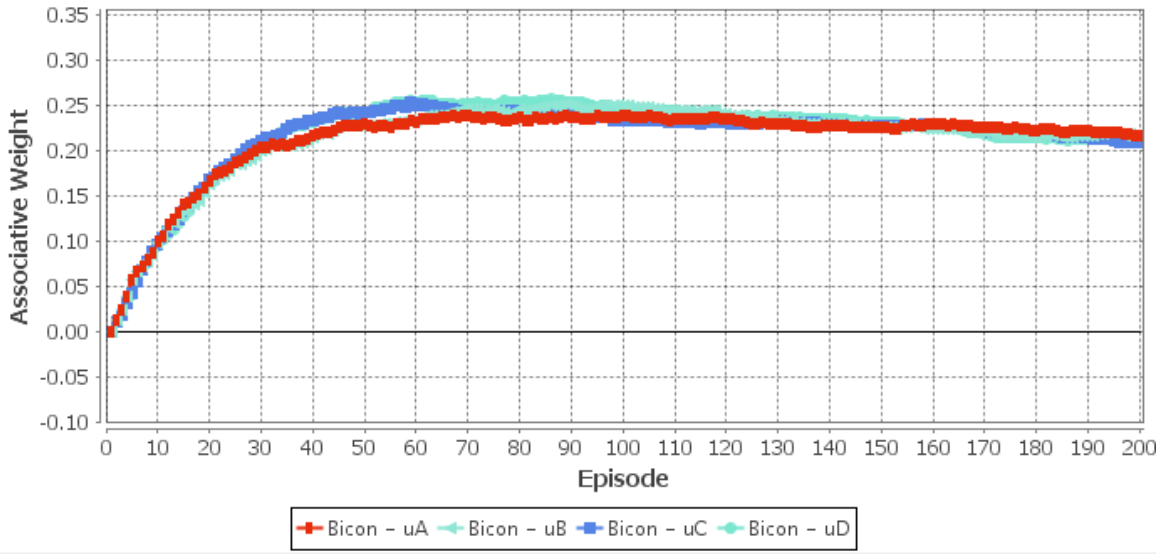


Fig. 3.60 Simulated weights of the unique elements of the CSs in the biconditional discrimination design with 10% common elements.

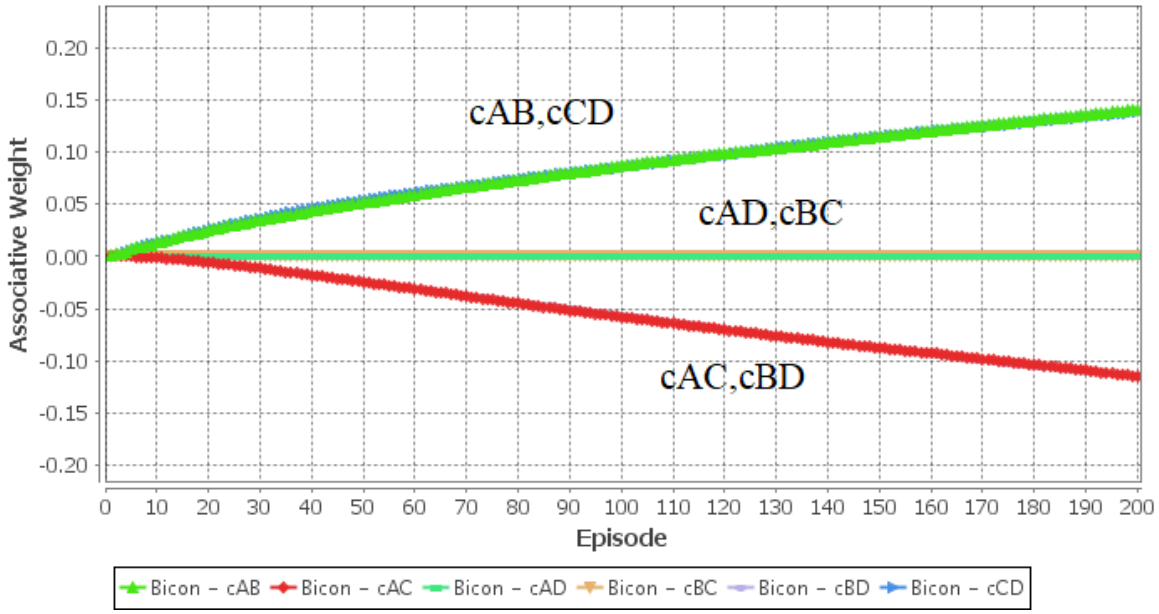


Fig. 3.61 Simulated weights of the common elements of the CSs in the biconditional discrimination design with 10% common elements.

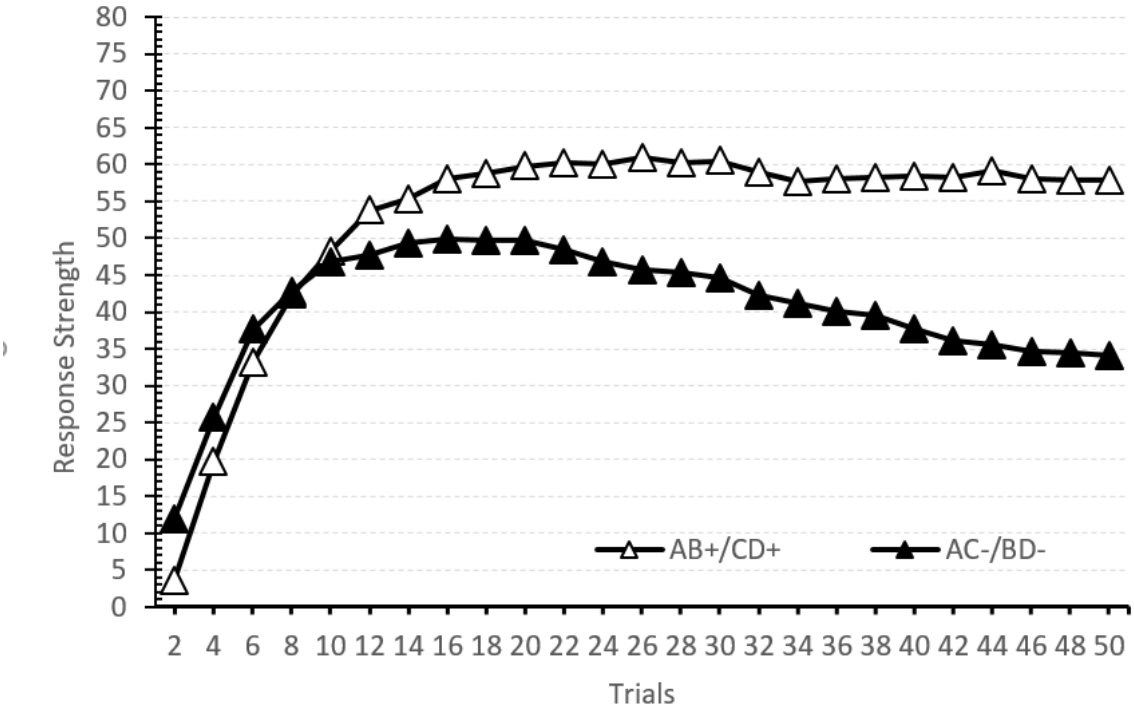


Fig. 3.62 Simulated response strength in the AB+/CD+ and AD-/BC- trials with 10% common elements. Average correlation $R = 0.91$.

Generalization Decrement Effects: Overshadowing and External Inhibition Asymmetry

A significant hurdle for configural models of learning is the observation that removing a cue from a previously reinforced compound stimulus (overshadowing) produces a larger decrement in responding than when a novel cue is added to the compound (external inhibition). As an elemental model, the DE model inherits an advantage in explaining this effect, in that it assumes a high degree of summation between the responding elicited by individual cues and their compounds. To demonstrate this, we will present a simulation of the aforesaid effect.

Precisely this difference between external inhibition and overshadowing was studied in (Brandon et al., 2000), using a rabbit eyeblink conditioning. In the experiment, Group A, AB, and ABC received respectively reinforced A, AB, and ABC trials. Subsequently all groups received test trials of randomly presented reinforced and non-reinforced presentations of A, AB, and ABC. The experiment found that adding a cue to a previously reinforced compound produced a smaller decrement in responding than removing a cue from a previously reinforced compound. Figure 3.64, left panel, the decrement from AB to ABC trials of Group AB is significantly smaller as compared to the decrement from ABC to AB of Group ABC.

Group	Phase 1	Phase 2
A	595*A*+	$24A+/24A-/24AB+/24AB-/24ABC+/24ABC-[random]$
AB	595*AB*+	$24A+/24A-/24AB+/24AB-/24ABC+/24ABC-[random]$
ABC	595*ABC*+	$24A+/24A-/24AB+/24AB-/24ABC+/24ABC-[random]$

Fig. 3.63 Brandon et. al. 2000 experimental design.

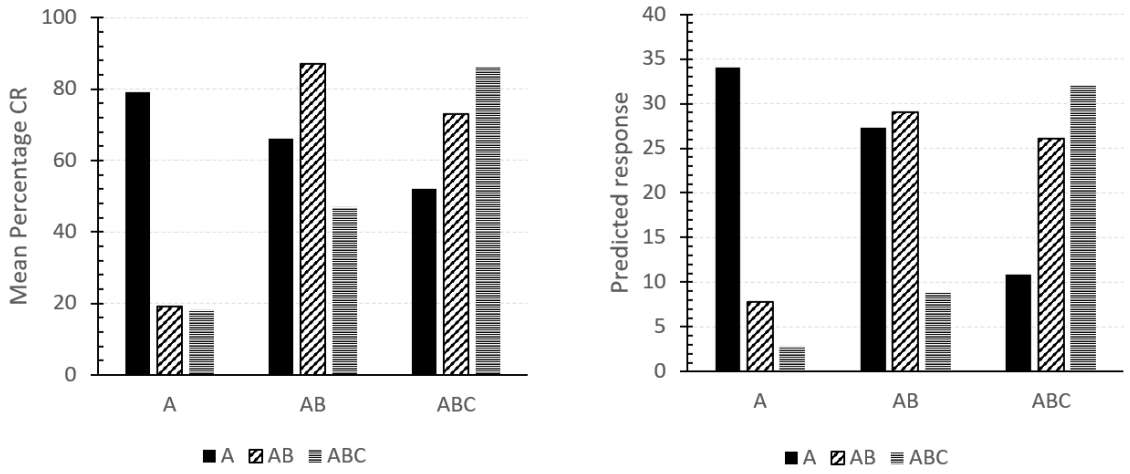


Fig. 3.64 Left: Brandon et al. responses in phase 2, Right: Simulation of Brandon et. al. 2000, responses in second phase.

The experiment was simulated using the following temporal parameters: a 10 time-unit CS, 1 time-unit US, and 100 time-unit ITI. The parameters were set per Table 3.1. The first phase groups A, AB, and ABC received respectively 20 programmed reinforced trials of type A, AB, and ABC. The second phase was programmed to present 20 randomly presented trials of types A, AB, and ABC. Each of these trial-types was reinforced on half of the presented trials. The full experimental design is in Table 3.63. The simulation reproduced this empirical ordering (Figure 3.64, right panel). Group AB displayed a lower generalization decrement from AB to ABC trials when juxtaposed to the drop in response strength from ABC to AB trials for Group ABC. The model produces such a result due to its elemental ontology leading to strong summation of constituent associative strengths of a compound. The decrement from AB to ABC in Group AB results from cue C rapidly becoming a conditioned inhibitor towards the US in the second phase due to the pre-existing AB to US strength from the previous phase and the novelty of C.

Configural Discrimination

In (Pearce, 1994), a discrimination of the form $A/B/C^{+}$ $AB/BC/AC^{+}$ ABC^{-} is discussed, and an argument is made that such a discrimination poses difficulties for elemental models. The experiment was conducted using an appetitive procedure in pigeons, and found that the pigeons learned to withhold responding on ABC trials (Figure 3.65). We intended to simulate this discrimination to test the limits of the DE model, as the model does not possess elements common to more than two stimuli in its stimulus representation. As such, one could expect that the model would be incapable of solving such a discrimination.

The simulation used a 1 time-unit CS, 1 time-unit US, and 50 time-unit ITI. The CS α_n starting value was 0.8 and the CS salience 0.2. The remaining parameters were set as per Table 3.1. The simulation was programmed to consist of 40 randomly ordered reinforced compound trials of the types A+, B+, C+, AB+, BC+, and AC+. Additionally, intermixed were 240 randomly ordered non-reinforced trials of compound ABC.

The model was able to reproduce the broad trends of learning. Responding on ABC trials was significantly lower than on other trials at the end of training (Figure 3.65). This resulted from the common elements of AB, AC, and BC gaining inhibitory strength towards the US (Figure 3.66, while the unique elements tended toward zero associative strength 3.67. Finally, responding on non-ABC trials was raised through the context gaining a large degree of associative strength (Figure 3.68). Thus, the DE model is in some instances capable of solving more complex discriminations than would be expected from simply the use of common elements.

Group	Phase 1
Discrim	$40A+/40B+/40C+/40AB+/40BC+/40AC+/240ABC-$ [random]

Table 3.15 Configural discrimination simulation design.

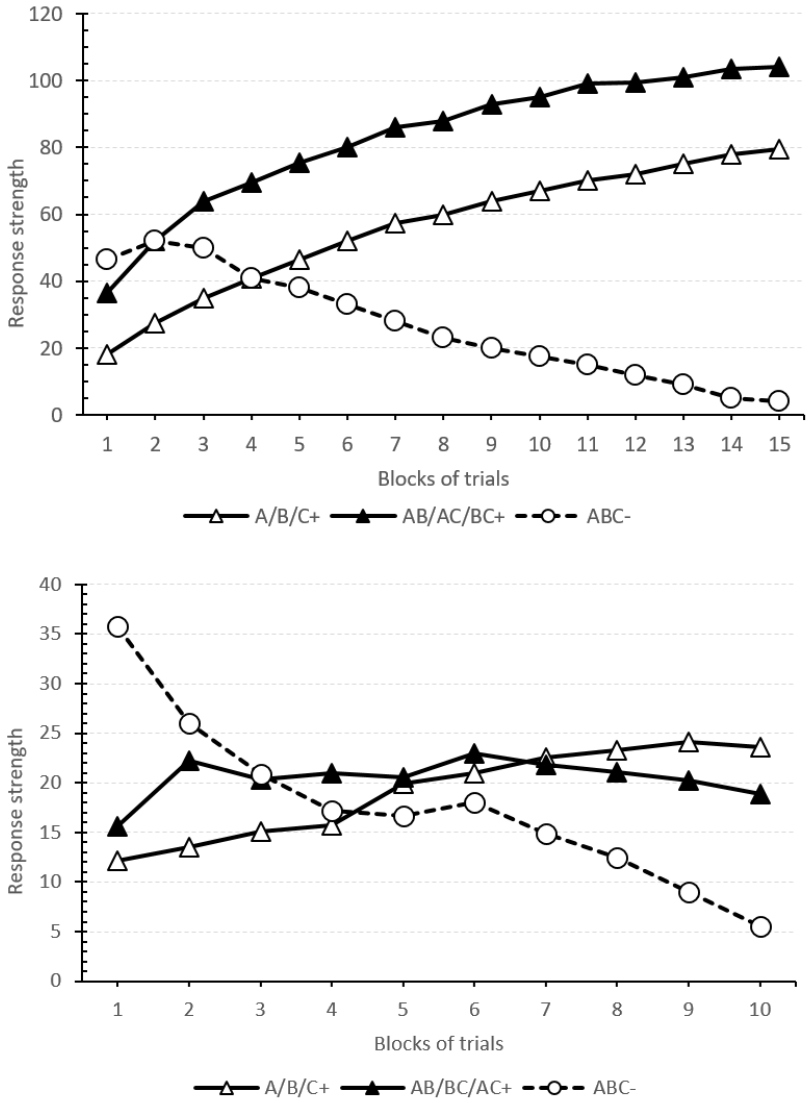


Fig. 3.65 Responses in the configural discrimination simulation.

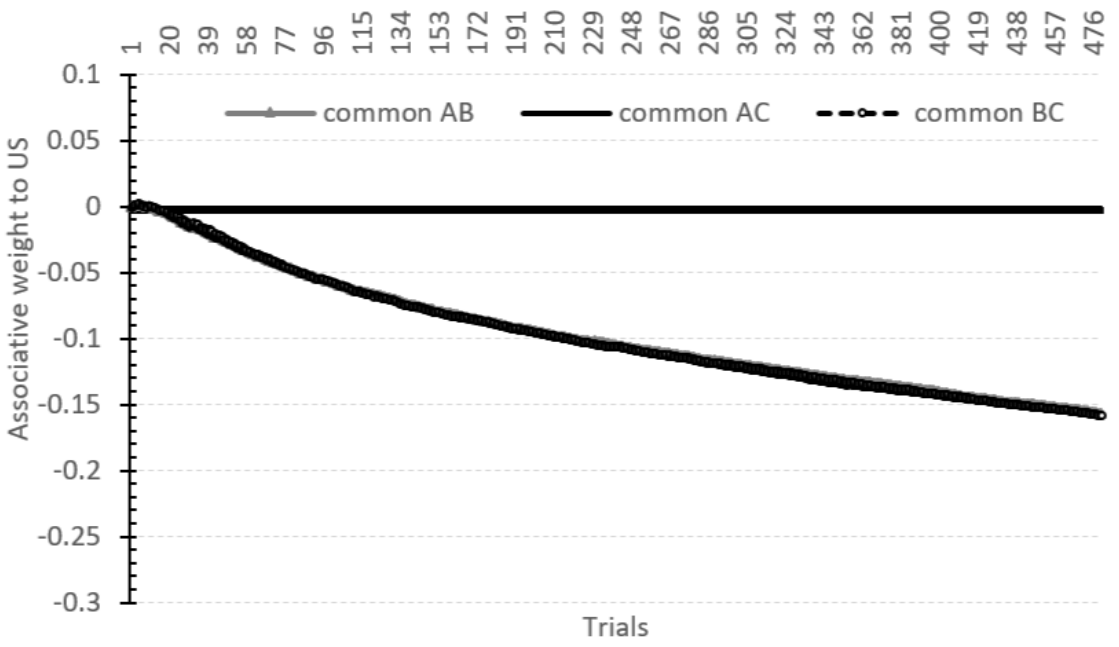


Fig. 3.66 Common element weights in the configural discrimination simulation.

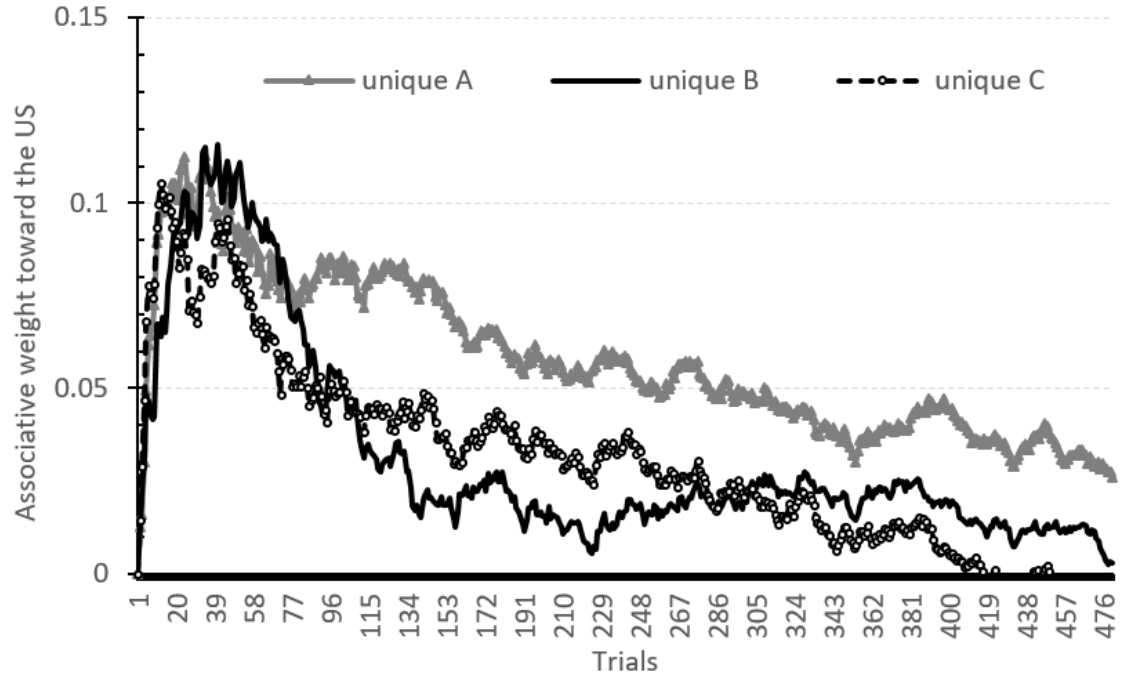


Fig. 3.67 Unique element weights in the configural discrimination simulation.

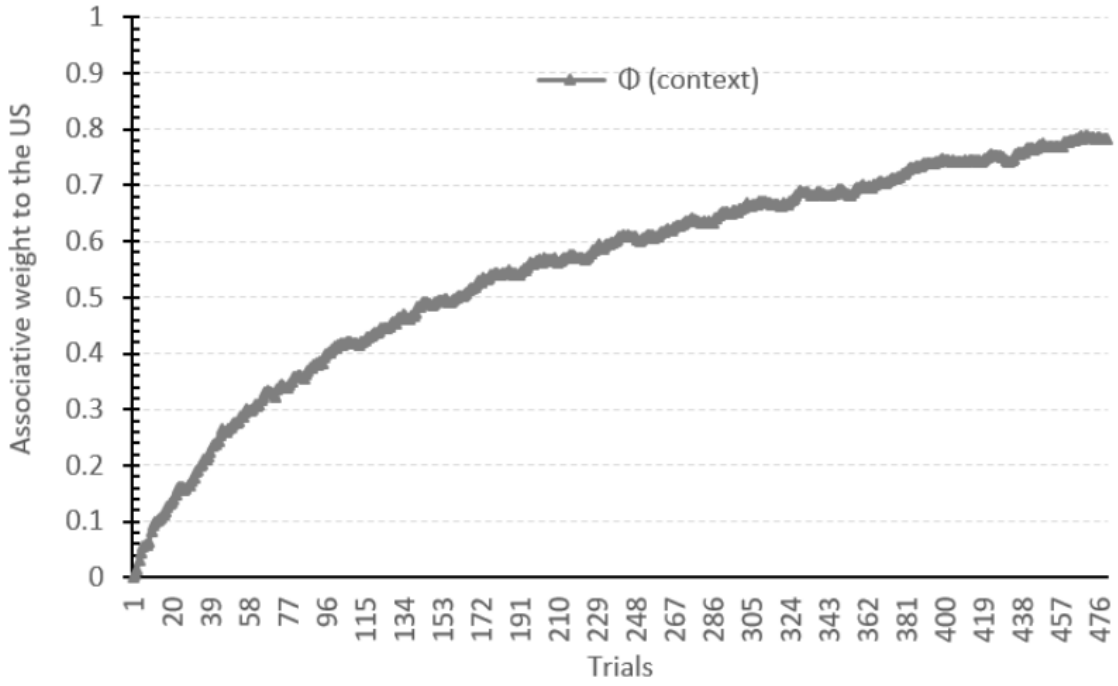


Fig. 3.68 Context weight in the configural discrimination simulation.

Timing Effects

Scalar Invariance i.e. Weber's Law

In the area of animal timing theory, a principle phenomenon is the scalar invariance effect. This effect is observed in that the anticipatory CR created through acquisition scales with a CS of a longer duration, such that the CR curve's variance and length stay in proportion (Gallistel and Gibbon, 2000). Hence, when these anticipatory CR curves are super-imposed, they display scale invariance. As the DE model utilizes semi-Gaussian curves which preserve the ratio between the standard deviation of the curve and its mean, the model is capable of producing approximate scalar invariance.

Group	Phase 1	Phase 2
Oversh.	$20A \rightarrow B \rightarrow +$	A?
Ctrl.	$20A \rightarrow +$	A?

Table 3.16 Scalar invariance simulation design.

The simulation was conducted using a 15 time-unit (Group 15) and 30 time-unit (Group 30) CS, 1 time-unit US, and 100 time-unit ITI. The CS and context saliences were respectively 0.2 and 0.1. The remaining parameters were set per Table 3.1. Each group received 50 reinforced trials. The full design is displayed in Figure 3.16.

The individual anticipatory CR curves as displayed in Figure 3.69. As can be seen, the curve for the 30 time-unit CS reaches a lower peak and has a wider spread. Of note is the minor inhibition occurring at the beginning of the anticipatory CR, which occurs due to cue-competition within the CS. A similar pattern of inhibition was found by (Mondragón et al., 1993), which demonstrates suppression ratios significantly exceeding 0.5. When both curves of the simulation are normalized by their response strength, they super-impose to a

large degree (Figure 3.70).

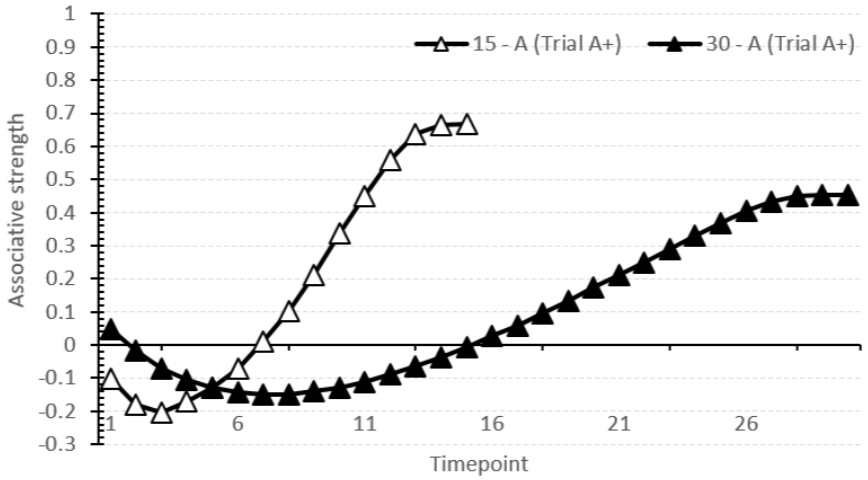


Fig. 3.69 Un-normalized anticipatory CR curves in the simulation of the scalar invariance effect.

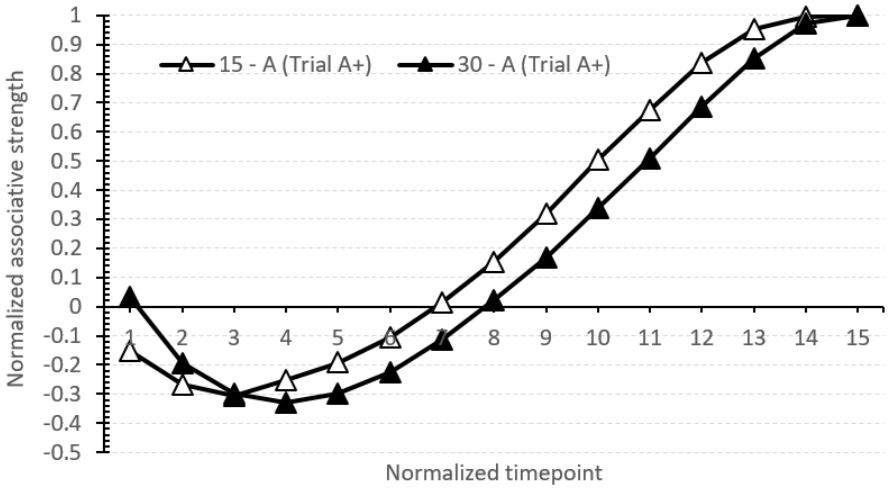


Fig. 3.70 Normalized anticipatory CR curves in the simulation of the scalar invariance effect.

Temporal Overshadowing

Temporal overshadowing is observed, when a CS B is introduced between the ISI of a CS A and US, such that CS B occurs after A and before the US. When compared to a control, this type of serial-compound presentation reduces the amount of learning accrued by stimulus A (Blaisdell et al., 1998).

The temporal parameters used for the simulation were a 10 time-unit CS, 1 time-unit US, 10 time-unit ISI, and 50 time-unit ITI. The remaining parameters were set according to Table 3.1. Group Oversh. was programmed to be presented with 20 trials of serial compound AB, followed by reinforcement. Group Ctrl. in contrast obtained 20 presentations of A followed by the US. This was followed by a test phase of cue A. In both groups, the period between the onset of A and the onset of the US was set equal. The full design of the simulation is in Table 3.17.

Group	Phase 1	Phase 2
Oversh.	$20A \rightarrow B \rightarrow +$	A?
Ctrl.	$20A \rightarrow +$	A?

Table 3.17 Temporal overshadowing design.

The simulation reproduced temporal overshadowing, with Group Oversh. displaying significantly less responding than the control (Figure 3.71). The effect is explained by the DE model as arising from the more proximate CS B in the Oversh. group obtaining a better activity overlap with the US representation during acquisition, which thence allows it to out-compete the CS A. This is borne out by the CS B in Group Oversh. having a larger weight toward the US in the beginning of the test phase than CS A (Figure 3.72).

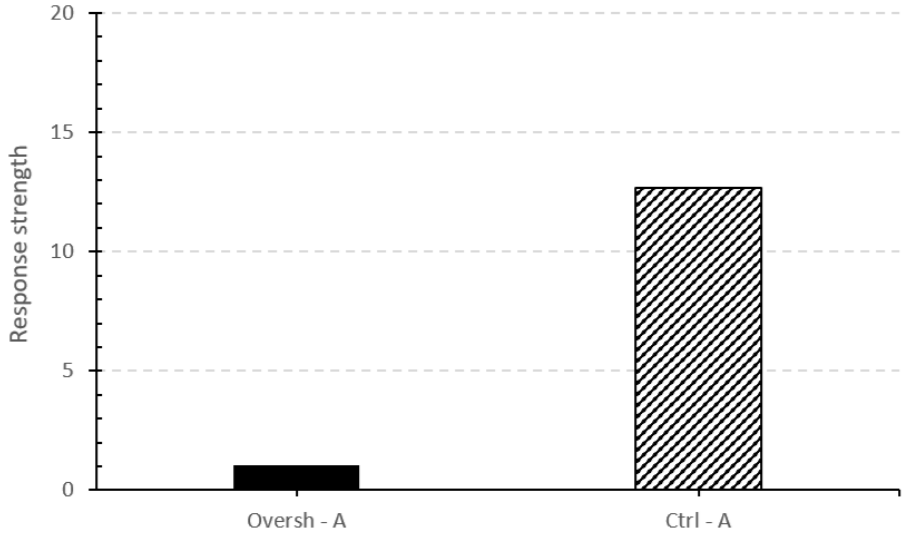


Fig. 3.71 Per-trial responding in the test phase of the temporal overshadowing simulation.

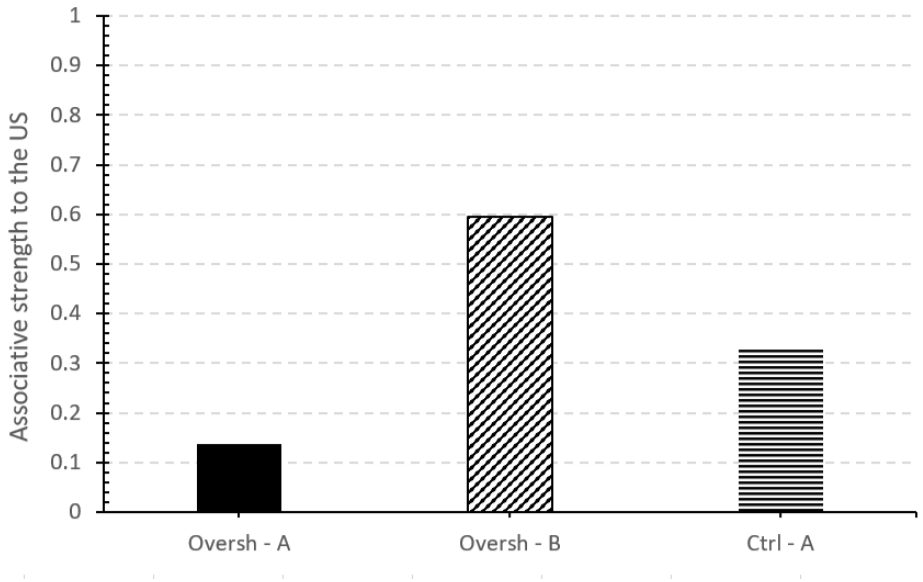


Fig. 3.72 Per-trial associative weights toward the US in the test phase of the temporal overshadowing simulation.

3.2 Discussion of Results

In the literature review we have discussed classes of phenomena, which are representative of psychological processes underlying animal learning, along with their relation to numerous prominent models in the field. These phenomena include standard acquisition, cue competition, pre-exposure, learning between neutral cues, mediated learning, configural discriminations, and timing effects. We have introduced a real-time computational model of associative learning, the Double Error (DE) model and demonstrated that it can reproduce these effects. The model's ontology is founded upon a connectionist network, wherein nodes representing the attributes of any stimulus (element), whether reinforced or not, enter into associative learning with one another in proportion to their state of activity. The defining algorithmic characteristics of the DE model are threefold. Firstly, it extends traditional error-correction learning by modulating learning by a predictor error-term, such that less novel predictors learn slower. Secondly, the model introduces a variable asymptote of learning that encapsulates principles of Hebbian neural principles of learning by postulating that elements with similar activity levels form stronger associations with one another. Thirdly, the model utilizes both a revaluation alpha towards reinforced and non-reinforced cues. These associabilities operate based on the principle that uncertainty in the occurrence of reinforced or non-reinforced are kept track of over a window of time through their moving average, and thence modulate the rate of learning of cues towards these classes of cues. Hence for instance, when the occurrence of reinforcement is uncertain, learning of associations towards reinforcers proceeds more rapidly. These associabilities further integrate a unique decay mechanism, which reduces the associability of cues towards reinforcing or non-reinforcing classes of cues, if the occurrence of these respective classes of is inherently random, such as in partial reinforcement. This decay operates based on whether a slower updating moving average of prediction errors crosses a threshold. Once this threshold is crossed, a decay

mechanism kicks in which reduces the associability progressively.

The DE model produces various acquisition and cue competition effects in a straightforward manner, stemming from its error-correction framework combined with the memory traces it uses for stimuli. Hence, shorter cues tend to acquire conditioning faster and to a higher asymptotic value than longer ones, due to their higher overlap with the activity of the US. Similarly, this activity overlap is most optimal when the CS and US co-occur, followed by when the CS occurs before the US. When the US occurs before the CS, the overlap is sub-optimal, and thus acquisition is less effective in this case (and inhibitory with a low value for the *b*/backward discount parameter of the model).

Acquisition is further affected by the ITI length, with longer ITIs leading to higher asymptotic associative link strengths between the CS and US due to context-US link extinction during the ITI. Acquisition is also affected by the exact parameters of the semi-Gaussian activation curves used, with a higher wave-constant parameter producing a more flat anticipatory CR. Finally, we have showed that partial reinforcement during acquisition leads to a CS forming an excitatory link to the US, which is approximately proportionate to the ratio of reinforced to non-reinforced trials in the procedure.

Next, we have simulated various cue-competition effects to display the predictive power of the error-correction framework utilized by the model. Conditioned inhibition and blocking follow from this error-correction framework equivalently to other models of this class. The phenomenon of unblocking however is of interest, as in this case the DE model postulates that the increased intensity or pro-longed exposure of a US can increase the level of conditioning to an otherwise blocked CS. With increased US intensity, this occurs through this intensity increasing the US activity, which hence increases the dynamic asymptote from

the blocked CS to the US. In the case of increased US exposure after the CS has occurred, an increased overlap between the activity of the blocked CS and the US lead to a longer period of excitatory learning. Finally, the last simulation of cue-competition involved the phenomenon of learned irrelevance, specifically an experiment displaying its absence. The DE model accounts for such learning by postulating that if the ratio of trials on which the CS is co-active with the US to those on which is active without the US is low, a context-mediated inhibitory CS-US link is produced. As such, this explanation differs from the account given by the Mackintosh model, which assumes that learned irrelevance arises as an attentional effect and as such does not predict this deviation from the classical result.

Similarly to the McLaren-Mackintosh, SLGK, and SOP models, learning between motivationally neutral cues allows the DE model to produce various 'silent learning' or pre-exposure effects. It predicts these effects as arising from both learning between neutral CSs, and through the constituent elements of a CS forming unitized nodes during pre-exposure. It postulates the same learning rules and processes both for reinforced and non-reinforced learning, thereby maintaining parsimony. That is, we maintain that reinforced outcomes should be treated as simply a subset of normal stimuli, as opposed to having completely unique ontology in terms of their stimulus representation. Its integration of these reinforced and non-reinforced alphas with its unique second error-term results in the model explaining pre-exposure effects as arising from the novelty of a cue decreasing as a result of learning. It predicts both the context specificity and sigmoidal shape of latent inhibition, with the latter resulting from the exponential changes in learning speed produced by the revaluation alpha. It hence uniquely assumes, due to these processes being dissociable, that increasing the uncertainty of reinforcers in general should attenuate this sigmoidal shape, while not completely removing the latent inhibition effect. The model further reproduces the effect observed by Leung et. al. whereby latent inhibition is enhanced by a pre-exposure of a CS compound. As

this effect is explained in the model by the cue undergoing reinforcement being predicted by the associative representation of the cue it was presented with prior, the implication is that presenting this retrieved cue in isolation before the acquisition phase should attenuate the observed effect. These pre-exposure related learning mechanisms also carry over to offering an account of the Hall-Pearce effect, by postulating that the reinforcement of a CS with a weak reinforcer acts in many ways as a form of pre-exposure treatment from the point of view of the model, due to the lower intensity of the US in the first phase treatment. Hence the model's explanation of this phenomena deviates from more traditional accounts like that of the Pearce-Hall model, which explains negative transfer as arising due to the CS fully predicting the US in the first phase, and thereby becoming less associable in later training. In contrast, the DE model assumes that the negative transfer effect will persist even with few acquisition trials with the weak outcome, as the decay in the associability of the CS is influenced more by the weak intensity of the US than by CS-US learning. Lastly, differential CS-CS learning and variable attention shifts between intermixed and blocked presentations of CS pre-exposure trials allow the model to account for perceptual learning being more pronounced with intermixed CS compounds.

The DE model's dynamic asymptote of learning produces either excitatory or inhibitory learning dependent on the causal connection between stimuli as tracked by the distance between the activations of the predicting and predicted stimulus. Mediated conditioning of a CS arises in the model through the combined effects of the aforesaid asymptote and associative retrieval by a second (previously paired) CS. Due to the asymmetry of the dynamic asymptote, learning between the associatively retrieved outcome of this CS and the US is excitatory and proportional to the degree to which the CS representation is retrieved. Hence the DE model assumes that the strength of mediated conditioning is proportional to the strength of within-compound associations, i.e. the length of CS-CS training. It is

additionally predicted that the observed direction of learning in mediated learning will vary based on prior learning towards the outcome. For instance, though the asymptote of the retrieved CS in backward blocking and mediated conditioning is the same, the direction of learning is claimed to be the opposite due to the initial associative strength of the retrieved cue differing in the designs. Further, the extent of backward blocking, we claim, should be proportional to the amount of reinforced compound trials in the first phase of a retrospective revaluation treatment. This prediction supersedes the static learning rules between A1 and A2 stimuli formulated in the extensions of SOP of Dickinson and Burke as well as Holland or the negative learning rate used in many models for mediated learning. The model rather proposes that learning between an absent and present cue is dependent upon how strongly the absent cue is retrieved, as well as on the prior link the absent cue has towards the present cue. This account of mediated learning occurs completely within the confines of an associative learning rule, thus not relying on non-associative mechanisms such as a comparator or re-playing of past experiences.

In terms of non-linear learning effects, the introduced elemental DE model can also reproduce a variety of configural discriminations without assuming that an animal has access to de novo configural information. It does so through a multi-factorial approach. The stimulus representation, which allows for elements common to multiple stimuli, produces effects such as negative patterning and biconditional discriminations similarly to the Rescorla-Wagner model with common elements. The DE model's revaluation alpha can amplify discriminatory learning through a unique decay process, which lowers the associability of cues presented with a persistently uncertain outcome. Its silent learning and attentional processes, as mentioned before, produce unitization of common and unique elements, as well as latent inhibition and loss of associability of common elements. In this latter aspect, it bears similarities to the CS-CS learning of the McLaren-Mackintosh and SLGK models.

It thereby explains the discrepancy between external inhibition and overshadowing without reference to additional representational structures such as the attentional buffer used in the Harris model. These common elements and learning processes also furnish an account of configural, non-linear discrimination effects, including negative patterning, biconditional discriminations, the configural discrimination, and perceptual learning using a completely elemental representation.

Finally, the DE model can account for some timing effects due to its real-time formulation. The model's semi-Gaussian activation time-waves allow it to produce approximately scalar invariant CR timing during acquisition. The phenomenon of temporal overshadowing is similarly accounted for, as presenting a competing CS between a CS and US leads to the CS closer to the US obtaining a more optimal activity overlap with the US representation, thus out-competing and overshadowing the more proximate CS.

Chapter 4

Mathematical Analysis of the Model

In this chapter we will further explore the mathematical properties of the model. We begin with the pseudo-code for the Double Error model algorithm. Next, we will demonstrate some convergence properties of the error-correction rule used in the model, to demonstrate that the learning in the model will not diverge to unexpected values, as well as to contrast the rate of learning to more traditional error-correction models. In the section thereafter, the relation of the double-error rule to measures of covariance is detailed.

List of abbreviations for Chapters 4-5

α/β	learning rate
δ	error
Δ	increment
e	Euler's constant
σ	standard deviation
σ^2	variance
$\sigma^2(a,b)$	covariance
ANN	Artificial neural network
BPTT	Backpropagation through time
CS	conditioned stimulus
DE	Double Error model
F	Functional similarity
H	Hidden layer
LSTM	Long short term memory RNN
MSE	Mean-squared error
P	prediction
PCA	Principal component analysis
RNN	Recurrent neural network
RW	Rescorla-Wagner model
TD	Temporal Difference model
US	unconditioned stimulus
w	associative weight
x	presence of input
Y	direct (feed-forward) activation
$\hat{Y}$	associative activation

4.1 Double Error Model Pseudo-code

Below is the pseudo-code for the entire DE algorithm. Of note is that the time and space complexity of the model is $O(n^2)$, due to weights being calculated at each time-point between each pair of elements. This is in contrast to US-centred models, such as RW, which scale linearly in time and space when the number of stimuli/elements is increased.

for $t = 1$ to T **do**
  for $p = 1$ to P **do**
    for $i = 1$ to I **do**
      $Y^t_{i,p} = exp(-\frac{(t-i)^2 s}{2c^2})$
      $\hat{Y}^t_{i,p} = \sum_m \sum_j w^t_{j,m \to i,p} \cdot \mathscr{A}^t_{j,m}$
      $\mathscr{A}^t_{i,p} = max(Y^t_{i,p}, \hat{Y}^t_{i,p})$
      $\delta^t_{US/CS} = (1 - \rho_r \mathscr{A}_{j,p})\delta^t_{US/CS} + \rho_r \mathscr{A}^t_{j,p} | \frac{\sum_o \sum_m \delta^t_{m,o}}{\#\ CS\ +\ US\ elements} |$
      $\hat{\delta}^t_{US/CS} = \begin{cases} \delta^t_{US/CS}, & if\ \delta^t_{US/CS} > \xi \\ 0, & otherwise \end{cases}$
      $\alpha_{r/n}(j,p)^t = (1 - \hat{\delta}^t_{US/CS})(1 - \rho_{r/n} \mathscr{A}^t_{j,p}) \alpha_{r/n}(j,p)^t + \rho_{r/n} \mathscr{A}^t_{j,p} |\delta^t_{US/CS}|$
      $\hat{\mathscr{A}}^t_{j,p} = \begin{cases} 1, & if\ Y^t_{j,p} > 0.1 \\ \mathscr{A}^t_{j,p}, & otherwise \end{cases}$
  for $p = 1$ to P **do**
    for $i = 1$ to I **do**
      for $o = 1$ to O **do**
        for $m = 1$ to M **do**
          $\lambda^t_{j,p \to m,o} = ||\hat{\mathscr{A}}^t_{j,p} \hat{\mathscr{A}}^t_{m,o}|| = \frac{\hat{\mathscr{A}}^t_{m,o} - |\hat{\mathscr{A}}^t_{m,o} - \hat{\mathscr{A}}^t_{j,p}|}{max(\hat{\mathscr{A}}^t_{m,o}, \hat{\mathscr{A}}^t_{j,p})}$
          $\delta^t_{j,p \to m,o} = \lambda^t_{j,p \to m,o} - \hat{Y}^t_{m,o}$
          $\delta^t_{j,p} = |\hat{\mathscr{A}}^t_{j,p} - \hat{Y}^t_{j,p}|$
          $\Delta w^t_{j,p \to m,o} = \delta^t_{j,p \to m,o} \cdot \delta^t_{j,p} \cdot s_{m,o} \cdot s_{j,p} \cdot \mathscr{A}^t_{j,p} \cdot \mathscr{A}^t_{m,o} \cdot \alpha(j,p)^t$

Algorithm 1 Double Error Model pseudo-code

4.2 Convergence Bounds of the model

Although the convergence results of a simple delta rule have been demonstrated analytically (Danks, 2003), such results do not apply directly for the Double Error learning rule due to the predictor error-term. Some properties of the learning equation can however be demonstrated mathematically. Therefore we will conduct such an analysis in this section, applied to an exemplary reduced case, to demonstrate the rate of learning in the model in comparison to the classical delta rule, as well as to show that the DE model does not produce divergent learning, i.e. is stable.

In the case of two stimuli (X and Y), which are completely active for an unlimited number of time-steps, an analytical calculation using a simplified version of the learning rule can demonstrate the rate of change as well as the lack of divergence of the weights in the model.

We will denote the weights $w^t_{X\to Y}$ and $w^t_{Y\to X}$ ($w^0_{X\to Y} = w^0_{Y\to X} = 0$), with activities $\mathscr{A}^t_X = 1$ and $\mathscr{A}^t_Y = 1$. Hence the simplified learning equations will look as follows:

$$w^{t+1}_{X\to Y} = w^t_{X\to Y} + \alpha(1 - w^t_{X\to Y}\cdot\mathscr{A}^t_X)|(1 - w^t_{Y\to X}\cdot\mathscr{A}^t_Y)| \tag{4.1}$$

$$w^{t+1}_{Y\to X} = w^t_{Y\to X} + \alpha(1 - w^t_{Y\to X}\cdot\mathscr{A}^t_Y)|(1 - w^t_{X\to Y}\cdot\mathscr{A}^t_X)| \tag{4.2}$$

Now, if α is sufficiently small, the absolute value of the predictor error-term will coincide with its non-absolute value for an arbitrary number of time-steps depending on α. Hence if we subtract one weight-update equation from the other, we see that:

$$w_{X\to Y}^{t+1} - w_{Y\to X}^{t+1} = w_{X\to Y}^{t} - w_{Y\to X}^{t} \tag{4.3}$$

That is, the distance between the two weights remains constant. Since both weights started from the same value, this means that they must in fact be equivalent. Therefore, we can write (as we assume that the activations equal 1):

$$w_{X\to Y}^{t+1} = w_{X\to Y}^{t} + \alpha(1 - w_{X\to Y}^{t})(1 - w_{X\to Y}^{t}) \tag{4.4}$$

Now, using the theory of generating functions (Wilf, 1990), we will expand this relation into a power series in the variable k. We will do so over the permissible values of the recurrence, $t \geq 0$:

$$\sum_{t=0}^{T} k^t w_{X\to Y}^{t+1} = \sum_{t=0}^{T} k^t (w_{X\to Y}^{t} + \alpha(1 - w_{X\to Y}^{t})(1 - w_{X\to Y}^{t})) \tag{4.5}$$

Next, we denote $B(k) = \sum_{t=0}^{T} k^t w_{X\to Y}^{t}$. Therefore, with simplification we have:

$$\frac{B(k) - w_{X\to Y}^{0} + k^{T+1} w_{X\to Y}^{T+1}}{k} = B(k) + \alpha \sum_{t=0}^{T} k^t (1 - w_{X\to Y}^{t})^2 \tag{4.6}$$

Now since we have $w_{X\to Y}^{0} = 0$, and grouping terms, we obtain:

$$k^T w_{X\to Y}^{T+1} = B(k)(1 - \frac{1}{k}) + \alpha \sum_{t=0}^{T} k^t (1 - w_{X\to Y}^{t})^2 \tag{4.7}$$

Thus, by setting $k = 1$ we see that the weight will grow in proportion to the square of $1 - w_{X\to Y}$ as expected. Thus the weight will reach asymptote considerably slower than

the standard delta rule. This is more obvious when we use the algebraic transformation of summation by parts with $k = 1$:

$$w_{X\to Y}^{T+1} = \alpha \sum_{t=0}^{T} (1 - w_{X\to Y}^{t})^2 = \alpha (\sum_{t=0}^{T} (1 - w_{X\to Y}^{t}) + \sum_{t=0}^{T-1} (w_{X\to Y}^{t} - w_{X\to Y}^{t+1}) \sum_{j=t}^{T} (1 - w_{X\to Y}^{t+1})) \tag{4.8}$$

As can be seen, the weight grows according to $\alpha \sum_{t=0}^{T} (1 - w_{X\to Y}^{t})$, plus the nested sum to its right. This nested sum is negative in an acquisition procedure, hence decrementing the first sum slightly. Since the first sum is proportional the the growth of an equivalent weight with the normal delta rule, this demonstrates that the weights using the double-error rule reach their asymptotic value slower. In fact, treating the recurrence as a differential equation produces the approximation $w_{X\to Y}^{t} \approx \frac{t}{t+1}$, as opposed to $w_{X\to Y}^{t} \approx 1 - e^{-t}$ for the standard delta rule. This difference is slight in practice. In Figure 4.1, a comparison between acquisition using the delta rule (Mondragón, E., Alonso, E., Fernández, A., and Gray, 2012) and the double-error rule was made. The simulation was conducted with 40 trials, CS $\alpha = 0.5$, and $\beta = 0.9$ in both models. The RW model converges to its asymptotic value slightly faster, as is visible in Figure 4.1. Of course, as the DE model due to its second error term is more self-correcting than a standard delta rule, its speed of learning could be substantially improved by using a higher alpha value without risking divergence of the associative links.

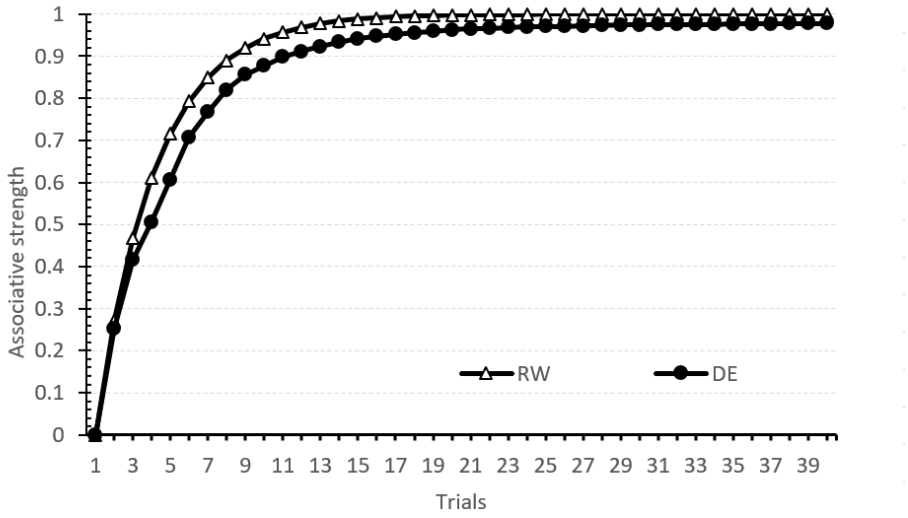


Fig. 4.1 Acquisition of CS-US association in the Rescorla-Wagner and Double Error models. Both simulations used 50 trials, CS $\alpha = 0.5$, $\beta = 0.9$.

The significance of this difference is that the double-error rule in effect has a stronger prior, requiring more evidence of a contingency to both acquire and extinguish an associative link between two stimuli representations.

Next we will show that the weights will not diverge under simplified conditions. First, for the case in which for some reason both weights have adapted themselves to a value of $1+\varepsilon$, the recurrence becomes (re-instating the absolute value):

$$w_{X\to Y}^{t+1} = 1+\varepsilon+\varepsilon(1-\alpha\varepsilon) \tag{4.9}$$

Since $1+\varepsilon(1-\alpha\varepsilon) < 1+\varepsilon$, the weights correctly tend towards lower values in this case. Similarly for the weights starting at $-1-\varepsilon$ we obtain:

$$w_{X\to Y}^{t+1} = -1-\varepsilon+4(1+\varepsilon)+\varepsilon^2 > -1-\varepsilon \tag{4.10}$$

Therefore, we have demonstrated the approximate growth rate of the learning equation in a simplified example scenario, as well as the tendency of the learning equation to be self-correcting under such conditions.

4.3 Double Error Learning and Covariance

The mathematical formulation of the Double Error model shares similarities with measures of covariance between variables. The covariance coefficient tracks the degree to which variation in one dimension correlates with variation in another dimension. The model can therefore be interpreted as producing predictions for the occurrence of a stimulus by learning how variations in the occurrence of the stimulus' representation correlate with variations in the representations of other stimuli. To demonstrate this connection, the Double Error learning rule will be derived from the unbiased covariance estimate equation.

The unbiased (based on whole population as opposed to sample) covariance estimate between element a and element b is given by the equation:

$$\sigma^2(a,b) = \frac{1}{N}\sum_{i=1}^{N}(x_{i,a} - \bar{x_a})(x_{i,b} - \bar{x_b}) \tag{4.11}$$

That is, for each example in the population or set we multiply the variation from the mean in one dimension, and multiply this by the variation in another. These individual values are then summed and divided by the number of examples or members of the population.

As a modification, the variable i will be interpreted as time instead of an example. Similarly, the mean of each variable will be re-interpreted as predictions for the occurrence of that variable. Predictions serve as momentary approximations of the mean value of a variable. The unbiased covariance estimate based on samples and the sample mean is hereby

transformed into a sequential covariance estimate utilizing a real-time estimate of the mean to calculate the variance.

$$\sigma^2(a,b) = \frac{1}{N}\sum_{t=1}^{N}(x_a^t - prediction(x_a^t|model))(x_b^t - prediction(x_b^t|model)) \tag{4.12}$$

The element x_a^t is further interpreted as the activation of the representation of stimulus a at time t, and the prediction as approximating the past activations of this stimulus. The two variances therefore become delta errors over time. The equation therefore now tracks the covariance of uncertainty. That is how deviation from the predicted activation value of the first stimulus correlates to deviation from the predicted activation value of the other stimulus:

$$\sigma^2(a,b) = \frac{1}{N}\sum_{t=1}^{N}\delta_a^t\delta_b^t \tag{4.13}$$

This equation can be further modified to incrementally approximate the correlation:

$$\sigma^2(a,b)^t \approx \sigma^2(a,b)^{t-1} + \alpha\delta_a^t\delta_b^t \tag{4.14}$$

The covariance can be used to calculate the slope m of the regression line (*i.e.* the line of best fit) $a = mb + c$ of the correlation between the two variables by dividing it by the variance of the first variable:

$$m = \frac{\sigma^2(a,b)}{\sigma^2(a)} \tag{4.15}$$

If the variance of a is equal to one, then the equation simply becomes:

$$m = \sigma^2(a,b) \approx \sigma^2(a,b)^{t-1} + \alpha\delta_a^t\delta_b^t \tag{4.16}$$

which is functionally similar to the Double Error associative strength update equation:

$$V_{(i,n)\to o}^t = V_{(i,n)\to o}^{t-1} + \mathscr{A}_{i,n}^{t-1}(\alpha_o\delta_o^t\alpha_n\delta_n^t) \tag{4.17}$$

The Double Error associative strength equation therefore calculates a measure similar to the slope ('m' in $b = m \cdot a + constant$ or $a = m \cdot b + constant$) of the regression line of two variables (stimuli). Due to this equality, it can be used to predict the relationship between the activity of one stimuli according to the activity of another. If we assume that stimuli activity ranges over all values from 0 to 1 (and there are time-points when both stimuli are active or inactive), then the axis-intercept or constant in the equation will tend to be minute. Thus, we arrive at the prediction equation:

$$P_o^t = \sum_n \mathscr{A}_n^t V_{n\to o}^t \approx \sum_n \mathscr{A}_n^t \sigma^2(o,n) \tag{4.18}$$

where the prediction for the outcome 'o' at time t is equal to the sum of predictions by individual stimuli, which are calculated as the product of a stimuli's activity at that time-point

and its associative strength to the outcome.

Lets now take a look at the TD associative strength equation (and transitively to Rescorla-Wagner):

$$V = \alpha\beta\,\delta^t x^t \tag{4.19}$$

It could be viewed as a similar measure of correlation if we interpret x^t as a delta error with a prediction of zero, ($\delta_x^t = x^t - 0$). Thus, the associative strength of the TD equation effectively calculates the dependency between the variance of the US from its mean, and the variance of x from a mean which has been shifted down to zero. Therefore we should expect that the TD algorithm should run into trouble when it is trying to predict the learning outcome of a situation in which the activity of x is lower than expected (lower than the mean activity, as estimated by the prediction), like during backward blocking. In backward blocking, the TD associative strength equation implies that no learning between the absent stimulus x and the outcome takes place. Yet, the associative strength between x and the outcome tends towards the negative. This can be seen as the slope of the regression line shifting its angle counter-clockwise due to a greater quantity of data-points in the quadrant of the coordinate plane where δ_{US} is positive and δ_x is negative. In fact this analysis predicts that the same behaviour should occur when the US is expected but absent and x is present. A corollary is that the TD and RW learning rules can be seen as improvements upon Hebbian learning of the form $V = y \cdot x$, where learning occurs to the extent that two stimuli are active at the same time. If both y and x are viewed as delta-errors with means shifted to zero like in the TD example, then Hebbian learning will effectively calculate a regression line slope, which is always positive. It therefore does not account for learning occurring when either one of the variables is active to a degree which is lower than expected.

4.4 Discussion

In this chapter, we have introduced the complete pseudo-code for the DE model and shown its time and space complexity to lie in the polynomial complexity class $O(n^2)$. Thus, extending the DE model to very large networks might necessitate that associations are not calculated between all elements, but rather follow principles of local connectivity. Next we demonstrated, using a simplified example and techniques of algebra and differential equations, that the double error of the model leads to slower learning than in the classical delta rule. This can be interpreted as the model requiring more evidence to modify learning (i.e. a stronger prior). Slower learning is not necessarily a weakness, as it to some degree ameliorates the problem of *catastrophic forgetting* (French, 1992) often encountered in neural networks, i.e. new learning erasing past learning. We also showed that the learning rule of the model will not diverge to infinite values in the scenario examined. In the section on covariance, we analysed how the model bears similarities to a dynamic covariance estimate between the occurrences of cues. This is of importance, as it implies that the learning of the model can be connected to a sound statistical basis. The analysis conducted is also the motivation for our assumption that the DE model can act as a general model of learning between correlated events, and as such is the starting point for the recurrent neural network extensions detailed in the future work section.

Chapter 5

Relation to Neural Networks

In this chapter, we will elaborate upon the connection between the mathematical equations behind the Double Error model and models of learning in general, and similar models in the field of machine learning. Though it might seem that there is considerable distance between the fields of classical learning theory and machine learning, in fact developments in both fields have often resulted from mutual theoretical cross-pollination (Hassabis et al., 2017).

We begin by detailing the relation between the predictor error-term of the model to backpropagation used to train feed-forward neural networks, as well as how normalizations of this process connect to the error-term. Next, a section in which a simplified version of the DE model is contrasted against a perceptron and a naive Bayes implementation on a contingency classification task is presented.

5.1 Relation of the predictor-error term to backpropagation processes

The relation, discussed above, of the Double Error associative strength equation to measures of covariance allows for a comparison of this learning equation with algorithms, such as backpropagation, used in the training of artificial neural networks. Backpropagation changes the weight w_i in a network by incrementally updating its value towards the negative of the derivative of the error E with respect to the individual weight:

$$\Delta w_i = -\alpha \frac{\partial E}{\partial w_i} \tag{5.1}$$

In the case of a squared error function $E = \frac{1}{2}(t-y)^2$, this becomes:

$$\Delta w_i = \alpha (t-y) \varphi' x_i \tag{5.2}$$

where t is the target, y is the current output of the network, φ' is the derivative of the activation function and x_i is the 'i'th input. Since φ' is used to assign portions of the error to network units based on their contribution, the equation can be simplified to:

$$\Delta w_i = \alpha (t-y) x_i \tag{5.3}$$

We assume that the value value t reflects the target output of this unit in the network (so its scaled by the derivative of the activation function). If x_i is viewed instead as a delta-error of $x_i - 0$ (i.e. with the mean shifted to zero), a similarity between backpropagation and the standard TD learning rule is visible. That is, both calculate covariance with a mean shifted to

zero for one of the variables. The implication is that backpropagation will display difficulty in learning a relation between one of the network units having a value below its mean, and the output of the network. That is, standard backpropagation does not incorporate a mean activation value to which the activation of a unit is compared for additional informational value.

One common optimization technique when using backpropagation is to normalize the inputs so the average value of each input is close to zero:

"Convergence is usually faster if the average of each input variable over the training set is close to zero. To see this, consider the extreme case where all the inputs are positive. Weights to a particular node in the first weight layer are updated by an amount proportional to δx where δ is the (scalar) error at that node and x is the input vector (see equations (1.5) and (1.10)). When all of the components of an input vector are positive, all of the updates of weights that feed into a node will be the same sign (i.e. sign(δ)). As a result, these weights can only all decrease or all increase together for a given input pattern. Thus, if a weight vector must change direction it can only do so by zigzagging which is inefficient and thus very slow." (LeCun et al., 2002, p. 16).

Normalizing the weights in this way is achieved by subtracting the approximate mean value of the input variable. Thus x_i becomes $(x_i - \bar{x_i})$, i.e. the simplified weight rule becomes:

$$\Delta w_i = \alpha(t-y)(x_i - \bar{x_i}) \tag{5.4}$$

which is similar to the Double Error rule if you view the target and output of the network as the current value and prediction of a 'predicted stimulus'. Similarly the mean subtracted

from the input variable x_i could be seen as the error for the predicting stimulus in the DE model.

Another optimization technique for artificial neural networks consists of input variable decorrelation:

"Inputs that are linearly dependent (the extreme case of correlation) may also produce degeneracies which may slow learning. Consider the case where one input is always twice the other input ($z_2 = 2z_1$). The network output is constant along lines $W_2 = v - (1/2)W_1$, where v is a constant. Thus, the gradient is zero along these directions (see Figure 1.2). Moving along these lines has absolutely no effect on learning. We are trying to solve in 2-D what is effectively only a 1-D problem." (LeCun et al., 2002, p. 17)

The usual approach is to use principal component analysis to remove linear dependencies from inputs. PCA describes the data through weighted covariance vectors, which are perpendicular to each other. One of these components will be the equivalent of the line of best fit, while others will describe spread in other dimensions.

If we view Double Error stimuli associative strengths as each mapping a covariance relation between two stimuli, then a combination of active normal and configural stimuli and their momentary temporal-element activations will form a linear combination of covariances, effectively weighted according to their activation. This prediction for a stimulus thus in essence forms a covariance decomposition of said predicted stimulus.

Take an example where stimuli A, B and X are predicting whether reinforcement occurs in A+/B+/ABX-. The associative strength of both A and B are equivalent to positive

covariance between these stimuli and the US (since values of A and B above their mean occur when the value of the US is above its mean value). Contrarily the stimulus X, which occurs on the non-reinforced trials, accrues an associative strength that corresponds to a negative covariance (activation above the mean value of X occurs when the US is below its mean value). Therefore after some training, the prediction for the occurrence of the US on non-reinforced trials will be a linear combination of the covariance relations of A,B and X with the US.

The Double Error model also has a relation to networks trained with multiple output units, for instance (Caruana, 1998, p. 164), which finds that "Multitask Learning is an inductive transfer method that improves generalization accuracy on a main task by using the information contained in the training signals of other related tasks. It does this by learning the extra tasks in parallel with the main task while using a shared representation".

The benefit of these networks is that training with examples not directly relevant to the primary output/task can still help train the network. Since in the DE model any stimulus can be both a predictor and a predicted stimulus, information regarding the presence of one stimulus will have implications for the associative strength between various other present or associatively activated stimuli. Therefore learning between neutral stimuli or stimuli and different USs can influence the pattern of conditioning observed.

Finally, the non-linearity dealt with hidden layers of a neural network are handled by compound stimuli or configural cues in classical conditioning models. In fact a strong equivalence exists between these two approaches, as configural ANN approaches such as Sigmi-Pi units (Courrieu, 2008) have been found to theoretically substitute for any quantity of hidden layers and units. Stimuli, compounds, and configurations can form direct associations with

the US. This implies TD and DE learning share similarities to a multilayer perceptron with short-cut connections from each layer to the output. These short-cut connections serve to allow linear portions of learning to be resolved through direct connections between nodes in a given layer and the output, while leaving the non-linear learning for higher layers in the network.

Therefore, to sum up, the DE model shares similarities to a multilayer perceptron trained with backpropagation, which incorporates short-cut connections between layers, incremental normalization through incremental covariance decomposition, and multi-task learning. This, along with the covariance analysis of the preceding section, acted as a motivation for fleshing out these similarities through both an instantiation of the DE model as a deep recurrent neural network with unsupervised learning capabilities, as well as using its second error-term to modulate the backpropagation of a vanilla RNN. These experiments are detailed in the following sections.

5.2 The Double Error Model as a Classifier

As a general model of learning, the DE model should be capable of learning the probability of one event given another. Therefore, in this section its ability to act as a classifier of the probability of an outcome given a configuration of inputs will be explored. To contrast its performance in this task, it is compared to two other models: naive Bayes and a single-layer perceptron.

The simulation of the models proceeds in two steps. First the parameters of the perceptron and DE models are optimized through a grid-search on the classification task involving respectively evaluation of 100 and 10 000 parameter values. The naive Bayes model used is parameter free. Next, the performance of each of the three models is compared using these optimal parameters.

Classification Task Specification

The classification task consists of the models predicting a number of randomly generated contingencies. Each contingency consists of a probability that a reward will arrive given a configuration of four inputs: X, A, B, and C. An example of such a sampled random contingency is presented in Table 5.1. The input X acts as a contextual node and is active on each 'trial'. The classification task consists of 10 randomly generated contingency sets, each of which is randomly and uniformly sampled for 1000 consecutive trials. Each contingency of 1000 trials is re-run 1000 times to obtain average performance of the model. The error measure used is the mean-squared error measured between the target value (0 or 1 depending on if the reward was present on a trial) and the output value of the model.

Contingency	Probability of outcome
X	0.52
A	0.19
B	0.63
C	0.07
AB	0.77
AC	0.35
BC	0.49
ABC	0.50

Table 5.1 Example of a random contingency sampled in the classifier task.

Model Specification

The simplified version of the DE model used consists of a set of 8 nodes in a connectionist network with weights bi-directionally between each two nodes. These nodes represent each configuration of possible inputs and one node represents the outcome. Each node has an associated activity level for a given trial. A simplified plan of the architecture is presented in Figure 5.1. The only free parameter of the reduced version of the model is α, the general learning rate. The predicted outcome class on a given trial is calculated using the dot-product of all nodes to the outcome. Prediction values above the bias are taken to indicate a prediction that the outcome will occur, while values equal to or below the bias are interpreted as a prediction the outcome will not occur. The equation used to calculate the prediction is as follows:

$$prediction = \begin{cases} outcome, & \sum_i x_i^t \cdot w_{i \to outcome}^t > bias \\ \neg outcome, & \sum_i x_i^t \cdot w_{i \to outcome}^t < bias \end{cases} \tag{5.5}$$

The activity of a given node is calculated as what ever is larger: its presence on a given trial (0 or 1), or the prediction for that node by other nodes (discounted by 0.5). The total prediction for one node is the dot-product from all other nodes to it. This value is confined to between -1 and 1. The error-terms are calculated equivalently to the full model. The weight update rule used is:

$$w_{i\to j}^{t+1} = w_{i\to j}^{t} + \alpha\delta_j^t\delta_i^t \tag{5.6}$$

The optimal alpha value found was $\alpha = 0.79$.

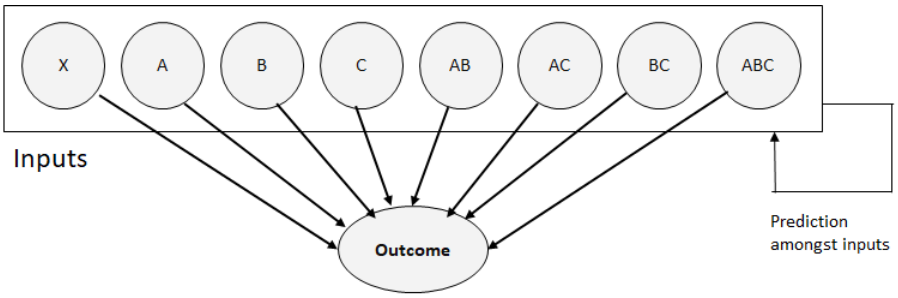


Fig. 5.1 DE classifier network.

For the perceptron, similarly eight nodes representing the outcome and input configurations were used (Figure 5.2). The free parameters are alpha and the bias of the perceptron. The activity of nodes is given by their presence on a given trial. The outcome was also calculated by whether the dot-product of weights to the outcome, plus the bias, was greater than zero:

$$prediction = \begin{cases} outcome, & \sum_i x_i^t \cdot w_{i\to outcome}^t > bias \\ \neg outcome, & \sum_i x_i^t \cdot w_{i\to outcome}^t < bias \end{cases} \tag{5.7}$$

The weights were updated according to the standard perceptron learning rule:

$$w_{i\to outcome}^{t+1} = w_{i\to outcome}^{t} + \alpha(target - output) \tag{5.8}$$

The optimal values found were $\alpha = 0.13$ and $bias = -0.8$.

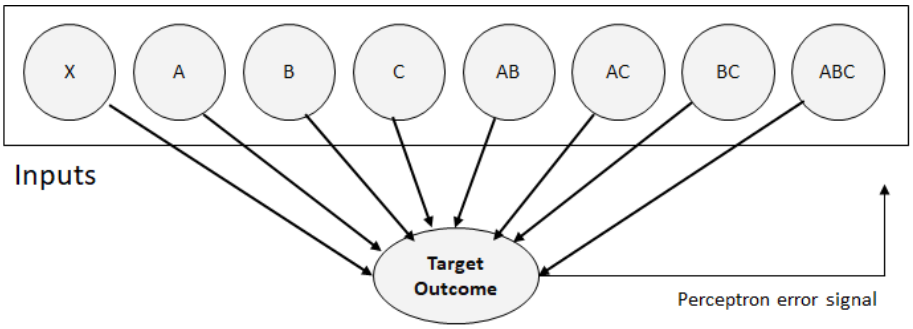


Fig. 5.2 Perceptron classifier network.

The naive Bayes model in return consisted of two sets of weights for each configuration, as well as presence values equivalent to the perceptron (Figure 5.3). The first set of weights tracks the number of times the outcome occurs in a configuration. The second calculates the number of times the configuration has occurred. The output of the model is a comparison of the probability of the outcome given the inputs present at a given trial, vs. the probability of the outcome not occurring given said inputs. It is calculated using the following equation:

$$prediction = \begin{cases} outcome, & \prod_i P(outcome|x_i) > \prod_i P(\neg outcome|x_i) \\ \neg outcome, & \prod_i P(outcome|x_i) \leq \prod_i P(\neg outcome|x_i) \end{cases} \tag{5.9}$$

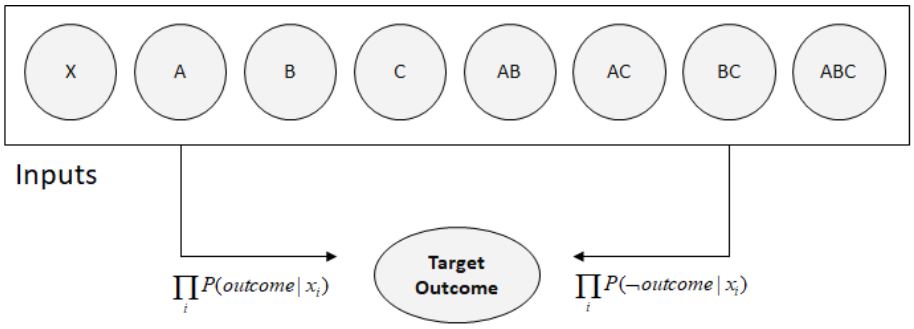


Fig. 5.3 Naive bayes classifier network.

Results

As is visible in Figure 5.4, all three models perform comparatively, with the DE model scoring a slightly lower MSE value on this task when compared to both other models, displaying that its rate of learning and asymptotic level of error on this task was roughly equivalent to other classifiers; i.e. the DE model is capable of performing as a classifier.

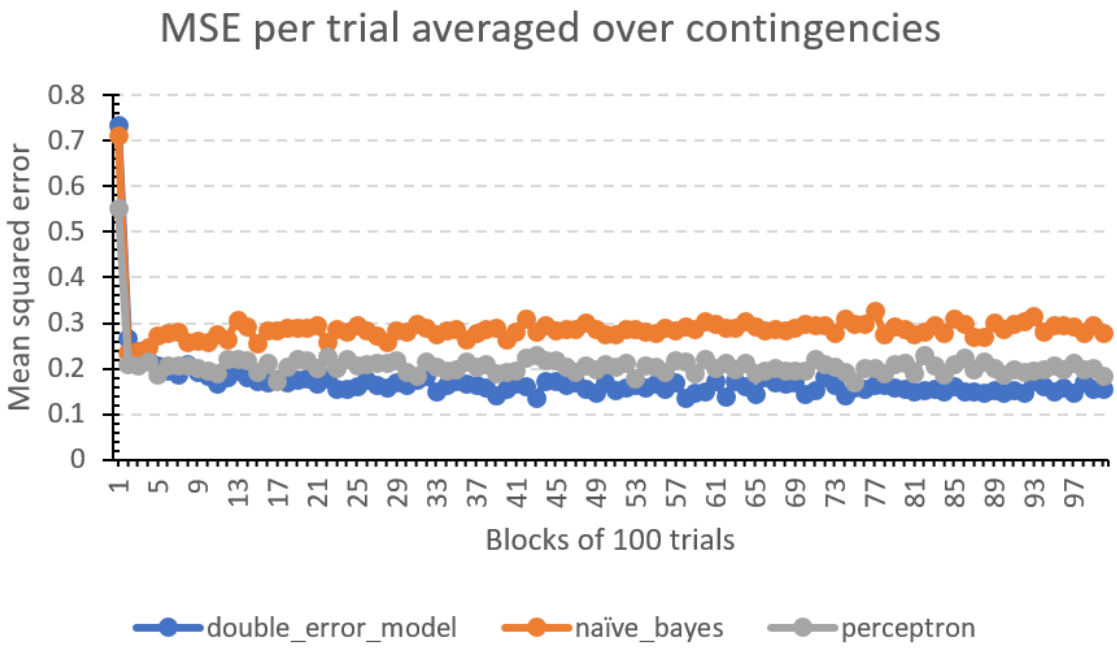


Fig. 5.4 MSE performance of DE model, naive Bayes, and perceptron on artificial classification task.

5.3 Relation to Recurrent Neural Networks and Machine Learning

In this section we will present some similarities of the DE model to other relevant models in machine learning theory. Firstly, commonalities of the model to reservoir networks and restricted Boltzmann machines will be elaborated. Next, similarities to recurrent networks will be discussed.

The DE model encapsulates many of the principles of reservoir networks (Schrauwen et al., 2007). Reservoir networks work on the principle that an input signal is fed to a reservoir consisting of randomly connected nodes making up a dynamical system, which then is sampled as a linear combination to produce a prediction for an output signal. Each of the nodes in this reservoir thus has a unique time and input-dependent activity, which is similar to the way that noisy time-waves with various means and standard deviations are utilized in the DE model to produce a prediction from a set of stimuli/elements to the outcome. In this sense, the stimulus acts as the input, its time-waves act like the reservoir, and the outcome (especially the US) acts as the error-signal. Specific types of reservoir nets have additionally been proposed to demonstrate principles of neurobiological plausibility, such as Liquid State Machines and its neuronal spiking process and random connections (Maass, 2010).

Next, the DE model encodes many principles used in restricted Boltzmann machines (RBMs), a type of generative stochastic neural network (Smolensky, 1987). RBMs, similarly to the DE model, have bi-directional modifiable links between nodes in the input and output layer of the network. These nodes usually take on binary values, and the energy function of a given RBM configuration is similar to that in a Hopfield network self-organizing map. Since the hidden-unit activations of an RBM are mutually independent and determined by the input,

the dynamics of the model share a resemblance to the way each element of the DE network predicts each other elements, and the total associative activation of an element is simply a linear factor of its incoming predictions. As I have demonstrated that the DE model's learning rule shares similarities with covariance estimates, there is a further resemblance in that RBMs in some sense perform factor analysis through the conditional probabilities of the input layer given the hidden layer (Cueto et al., 2010).

Processes of the DE model additionally share resemblance to many processes integral to classical recurrent neural networks (RNN) as well as more complicated RNNs such as Long Short-term Memory (LSTM). The fundamental premise of a recurrent neural network is that units of the network maintain a persistent memory of prior events through recurrent connections to themselves. In a basic recurrent neural network, this works through a simple recurrent connections from the node to itself (Figure 5.5). In more sophisticated networks, such as LSTMs (Hochreiter and Schmidhuber, 1997), the persistent memory state is maintained through the combination of a recurrent connection and a forget gate, which can be triggered by other units (Figure 5.6). This forget-gate behaviour is learned through backpropagation as well, hence allowing complex temporal dynamics to be learnt using the error gradient.

The principle of recurrence is widely utilized in connectionist models of learning, and forms the basis for models predicting conditioning wherein the offset of a CS occurs before the onset of a reward. This usually takes the form of either the persistence of the direct sensory stimulation of a stimulus, or through an eligibility trace, which is triggered by the onset of said stimulus. In this sense, models of conditioning can be said to already encapsulate some principles of recurrent networks. For instance, in the current DE model, recurrence occurs not only through the persistent memory trace initiated by the direct occurrence of a

stimulus, but also through associative retrieval re-activating the representation of a stimulus, making it eligible for further (mediated) learning, hence there is, like in LSTM networks, control over the contribution of a node to the overall prediction of an outcome, by other nodes it has previously been paired with through associative learning (instead of backpropagation).

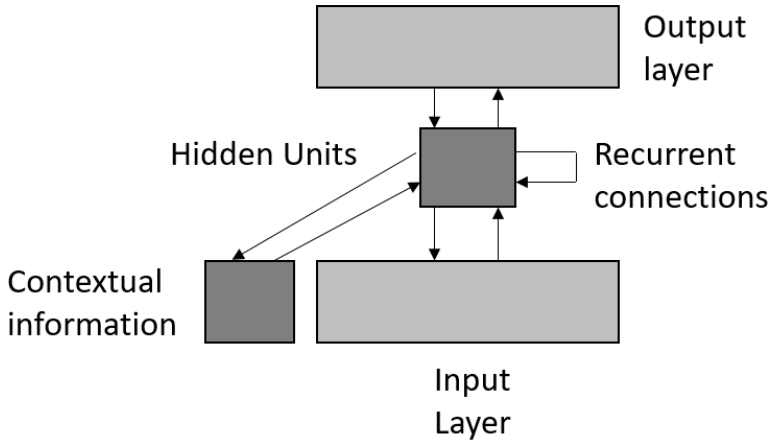


Fig. 5.5 RNN simplified architecture.

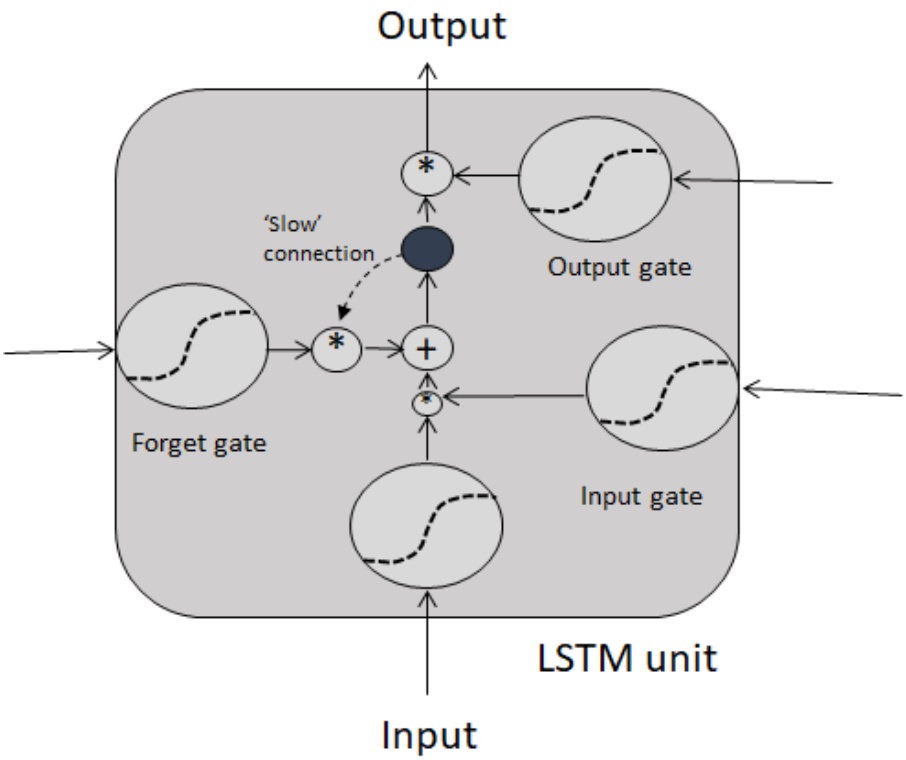


Fig. 5.6 LSTM simplified architecture.

5.4 Discussion

In this chapter we have demonstrated that the predictor error-term, associative retrieval, and dynamic asymptote of the DE model share strong similarities to extant concepts in the machine learning field. It shares similarities to a perceptron trained with backpropagation and normalization of such learning through incremental covariance decomposition. We have also postulated a connection to multi-task learning. As such, aspects of the DE model can be utilized in various tasks in this field. We have applied versions of the model to a generated contingency classification task and demonstrated performance competitive to two simple models, indicating that associative predictions between inputs in a connectionist network can improve classification performance in this task.

We believe some principles of the Double Error model could thus carry over to machine learning. For instance, the mediated learning of the model enables a principled manner of updating past learning during on-line learning. Additionally, updating past learning in this manner would tend to require fewer training samples for the algorithm to learn dependencies in the training samples. Combined with the memory traces of nodes, which persist differentially over long periods of time according to a semi-Gaussian activity curve, this could offer an alternative or complement to the computationally expensive Backpropagation Through Time (BPTT) weight-update algorithm. The second error-term of the model similarly modulates the total learning such that more novel inputs are preferentially associated. This makes statistical sense, as inputs that have occurred prior are less likely to be causally linked with novel inputs. Finally, the variable associability of nodes, which is divided into task-specific associability (in this case reinforced cues) as well as neutral associability (non-reinforced cues) and takes into account a long time-window of prediction errors, could potentially improve model performance by speeding up learning when these classes of nodes are uncertain,

i.e. when there is more to learn.

Chapter 6

General Discussion

In this thesis we have introduced a real time error-correction model, the Double Error model. It extends concepts of learning used in both the field of psychological learning theory and machine learning. Its main assumption is that the representations of events can be broken down into elements or nodes representing attributes of said events which are connected to one-another in a network through adaptive weights. These elements follow a distinct cascading activation pattern governed by probabilistic sampling guided by an activation equation using semi-Gaussian curves. It is further assumed that associative connections form between all such nodes irrespective of whether they are related to an outcome or not. The error-correction learning between these nodes incorporates both a novel variable asymptote of learning measuring the activity-correlation of nodes (instantiating Hebbian principles of learning), as well as a second, predictor error-term that modulates the speed of learning. Nodes that have previously formed associative links are subsequently capable of retrieving nodes that they have been paired with. This process of associative retrieval, along with the aforesaid asymptote, allows the Double Error model of accounting for a variety of learning phenomena wherein absent yet retrieved cues undergo changes in their associative strength towards other cues. The model further introduces an attentional process governed by a revaluation alpha, which changes in accordance to the general uncertainty of reinforced and

non-reinforced classes of cues. This revaluation alpha crucially incorporates a mechanism by which persistent uncertainty in the occurrence of cues (e.g. partial reinforcement) is re-evaluated as a source of information in itself, leading to a decline in attention. The introduced alphas hence raise the attention paid during periods of uncertainty in the occurrence of cues, and in some cases produce emergent effects in real time, such as a better predictor of the outcome gaining more attention. As shown in Chapter 3, as a general model of learning, DE accounts for phenomena predicted by extant learning models, and further uniquely predicts for instance the apparent contradiction of mediated learning in backward blocking and mediated conditioning proceeding in opposite directions. It is able to reproduce a wide variety of effects including general conditioning, cue-competition, pre-exposure effects, mediated learning, configural discriminations, as well as timing effects. The model was evaluated both through the juxtaposition of simulations and experimental learning data, demonstrating its considerable predictive power in the field of learning theory.

In Chapter 4 we have analysed the mathematical basis of the model. We indicated its complexity class ($O(n^2)$), and discussed how this might necessitate local connectivity adjustments for larger connectionist networks. Similarly, we showed that learning does not diverge in the model and is generally slower than with a classic delta rule. This can be interpreted as the model having a stronger prior than traditional delta rules, which may help protect it from catastrophic forgetting.

Finally, in Chapter 5, we analysed the DE model's similarity to processes governing backpropagation in a perceptron. This motivated the application of the DE model's processes of the predictor-error term, associative retrieval, and dynamic asymptote to classification tasks.

The contributions of the introduced DE model, detailed in this thesis, to the fields of learning theory and machine learning have been as follows:

- The DE model introduces a unique second, predictor error-term. This error-term attenuates the speed of learning between a predictor and outcome according to the extent to which the predictor is predicted by other cues and itself. This mechanism is integral to the model's ability to account for various pre-exposure effects. In Chapter 4 we have demonstrated the connection of this second error-term to covariances estimates, and in Chapter 5 we have shown that its use as a simple classifier.

- The model introduces a unique Hebbian-inspired dynamic asymptote of learning, which measures the similarity of activation levels between two elements. Thus, elements with more similar activity levels are capable of forming stronger associations with one another. This in a sense measures the likelihood that the two elements are causally linked, and endows the model with the ability to predict a wide range of learning effects, including mediated learning. Of specific note is that it can account for both backward blocking and mediated conditioning/SPC simultaneously.

- The model introduces revaluation alphas, which track a time-window of uncertainty in the occurrence of reinforced and non-reinforced stimuli. In addition, persistent uncertainty is turned into a source of information in and of itself. This is done by measuring whether the moving-average of stimulus uncertainty is over a threshold, and if so reducing attention to cues of which the occurrence is inherently random. These alphas allow the model to account for the sigmoidal shape of latent inhibition, in addition to numerous other effects.

The DE model similarly, as detailed in Chapter 2 and 3, produces a variety of distinct predictions in the field of learning theory. Some are novel to DE; others, although present to various degrees in other models, arise naturally in DE as a coherent corpus.

- For CSs of equal salience and sufficiently long duration, the best predictor of an outcome will capture the highest amount of attention.
- Associations form between any two stimuli, of which the representations are concurrently active.
- Retrospective correlation learning in learned irrelevance is mediated through CS and context-dependent activation of the US representation. Hence, stronger context-US links will produce a stronger effect.
- The sigmoidal shape of latent inhibition is proportional to the attenuation of selective attention paid to the CS, and hence proportional to the length of CS exposure.
- The context specificity of latent inhibition is produced by context to CS associations forming during pre-exposure, and is attenuated by exposing the context in isolation after CS exposure.
- The Hall-Pearce effect arises in part due to the weak-shock trials reducing the selective attention paid to the CS, as well as due to the formation of context-CS associations. As such, conducting the second phase of the treatment in a novel context should significantly attenuate the observed effect.
- The difference in perceptual learning between intermixed and blocked presentations of cues arises in part due to CS-CS association between the reinforced cue and its intermixed associate in the first phase being weaker than the CS-CS association between the reinforced cue and the CS presented in the block of trials. This leads to lower

mediated conditioning during the subsequent reinforced trials in the former condition. Hence, preceding the reinforced trials with non-reinforced trials of the subsequently reinforced CS should lead to a smaller difference in perceptual learning between the intermixed and blocked conditions.

- The DE model predicts that a CS, which was exposed in compound with another CS, experiences more latent inhibition than a CS exposed in isolation due to the former condition leading to the subsequently reinforced CS retrieving a representation of its associate. This retrieved associate in return produces a prediction for the retriever, reducing the retriever's speed of association toward the reinforcer. Hence, a further prediction is that should the retriever CS undergo further non-reinforced exposure after its exposure with its associate, the increased latent inhibition effect should be attenuated.

- The DE model predicts that mediated learning effects as a whole can be accounted for as the result of the combination of associative retrieval through within-compound associations and learning proceeding according to the similarity in activation between the representations of two stimuli (whether directly present or retrieved).

- Further, the DE model predicts that the apparent contradiction in learning proceeding in opposite directions in backward blocking and mediated conditioning/SPC procedures arises due to the prior training endowing the retrieved cue with respectively moderate or no associative strength. Hence, since the asymptote of learning is respectively lower and higher during the mediated learning, extinction and acquisition is observed.

- As the temporal overlap between cues influences the asymptote of learning for associations forming between them, the model additionally predicts that backward blocking is respectively facilitated and attenuated by more phase 1 training according to whether the CS used is of long or short duration.

- The model predicts that in general mediated conditioning, SPC, and BSP will occur in proportion to the extent to which the retrieved cue is retrieved. As such, more Phase 1 training should strengthen the observed effect.

- The DE model predicts the configural discrimination in Chapter 3 through the context becoming highly excitatory. Therefore, this hypothesis is falsifiable by measuring responding in a further context test phase.

- In terms of timing, the model predicts that the early portions of a sufficiently long CS will tend to become inhibitory toward the US after sufficient training during an acquisition procedure.

Finally, the introduced mechanisms of the DE model relate to and in some cases supersede learning mechanisms instantiated by existing models of associative learning:

- The DE model uses the general error-correction framework familiar from models such as RW, SB, and TD.

- Unlike TD, learning aims to establish a covariance between the co-occurrence of element activity. Hence, no temporal difference term is utilized.

- The time-waves used in the DE model bear a strong similarity to those introduced by the micro-stimulus representation of TD. However, the amplitude of time-waves are not normalized in the same way as with micro-stimuli.

- Unlike the REM and Harris models, the DE model does not propose that elements are differentially sampled or active depending on their context. Rather, common elements exert their effect in the DE model through their redundancy on compound trials (i.e. the same common element can not be sampled twice at the same time).

- The ontology of the DE model is completely elemental and does not involve explicit hidden layers (though the predictor error-term induces implied functional layering), and as such is at odds with the configural representations or hidden layers of for instance the Pearce, APECS, and SLGK models.

- The revaluation alphas of the DE model share strong similarities to the PH alpha rule, which similarly proposes that uncertainty raises selective attention. However the DE model diverges from the PH rule in that it predicts that prolonged uncertainty becomes a source of information, and that attention to reinforced and non-reinforced classes of cues involve separate attentional processes. Though the Mackintosh model's attentional rule seems opposed to that used in the DE model, the revaluation alphas can in fact produce effects whereby the best predictor of an outcome captures the most selective attention. This, however, is dependent upon sufficiently long CSs being used.

- The DE model has some similarities to the SLGK, SOP, and McLaren-Mackintosh model, in that all four models assume that the novelty of a stimulus influences the speed with which it forms associations. This leads to all four models predicting that key components of latent inhibition are CS unitization and context-CS links. Hence, the DE model is at odds with the assumption of the Hall-Rodriguez model that latent inhibition involves CS-noUS type learning.

- Unlike the McLaren-Mackintosh model, the DE model does not use a weight-decay mechanism, though we intend to attempt integrating such a mechanism in future work.

- The DE model does not use activation states as seen in SOP, instead learning proceeds according to the overall activation of an element, which is influenced both by its sensory inputs as well as predictions by other elements.

- Finally, in contrast to the replayed experience model of (Ludvig et al., 2017), the DE model accounts for mediated learning through associative retrieval and its dynamic asymptote.

Future Work

We have identified various future directions of work from the theory we have developed for the DE model. These involve both improvements and refinements upon the existing model, work attempting to tie the DE model and other models of associative to learning to Bayesian graphical models, a proposal to develop many of the underlying principles of the model into a deep connectionist model of classical conditioning, as well as two proof-of-concept experiments on applying parts of the DE model to NLP tasks through recurrent network implementations.

In terms of improvements to the extant structure of the model, two potential modifications have come to our attention. The first is to improve the temporal dynamics of the DE model to allow for more precise timing, and to connect such a process to current knowledge of neural learning in the brain. It has been shown that the exact timing of neuronal spiking is implicated in determining whether excitatory or inhibitory long term plasticity (LTP) is observed (Kobayashi and Poo, 2004). That is, the optimal timing for producing excitatory learning tends to involve one neuron firing slightly before the other. In contrast, if a neuron A fires before neuron B, this tends to weaken excitatory connections. Encoding such synaptic timing in the DE model would allow for a more realistic process of learning, and would replace the current b parameter that controls backward learning.

The second potential modification of the model is to implement multiple time-scale changes in associative weights. Such a process is used for instance by the McLaren-Mackintosh model (McLaren and Mackintosh, 2000), and allows for it to predict the Espinet effect. We believe such differential short and long-term learning dynamics could underlie some basic forms of reasoning and interact with the mediated learning processes of the DE model to account for effects such as higher-order mediated learning. Neural evidence exists

that such a process of multiple-time-scale LTP takes place in the brain through a combination of calcium signalling through NMDA and AMPA receptor dependent and direct axonal stimulation, as well as through long-term modifications of axonal structure and receptor expression (Barria et al., 1997; Lu et al., 2001).

Next, we are working on a project to explore how deep neural networks can serve as a way of directly encoding the stimulus representation used in classical conditioning models, instead of relying on abstractions with no direct connection to sensory data (Mondragón et al., 2017). We believe that learning stimulus representations directly from data would allow for more realistic (in terms of processes actually taking place in the brain) representations of how animals encode stimulus similarity and dependencies (and maintain connective sparsity), thus possibly shedding light on how complex discriminations are solved. Such a rich representation, based on multiple levels of abstraction emanating from the deep architecture of such a model, would then feed into associative learning mechanisms, and would possibly involve useful concepts from extant machine learning theory (e.g. recurrent connections).

In the next section, I will elaborate upon some of the brain-storming I have done to connect the theory of the DE model more strongly to Bayesian concepts, specifically Bayesian graphical models.

Relation to Bayesian Graphical Models

In graphical models of Bayesian networks, *support* denotes the log-likelihood ratio $log(\frac{P(d|Graph1}{P(d|Graph2)})$ between two (or more) competing graphical models which are being compared as likely candidates for explaining a data vector d. For instance, when comparing the log-likelihood that an effect is caused by one cause vs. two causes (Figure 6.1 left and right side, respectively), where w are parameters signifying the weighting of the effect. Support has been found to follow trends in human performance in tasks involving causal learning (Figure 3.4 in (Griffiths et al., 2008)). As such, this section aims to show that in a classical conditioning model incorporating error-correction and stimulus competition, the associative weight of a stimulus representation towards an outcome is similar to a measure of Bayesian support. The motivation is to strengthen the case that the Double Error model instantiated a form of approximate Bayesian reasoning.

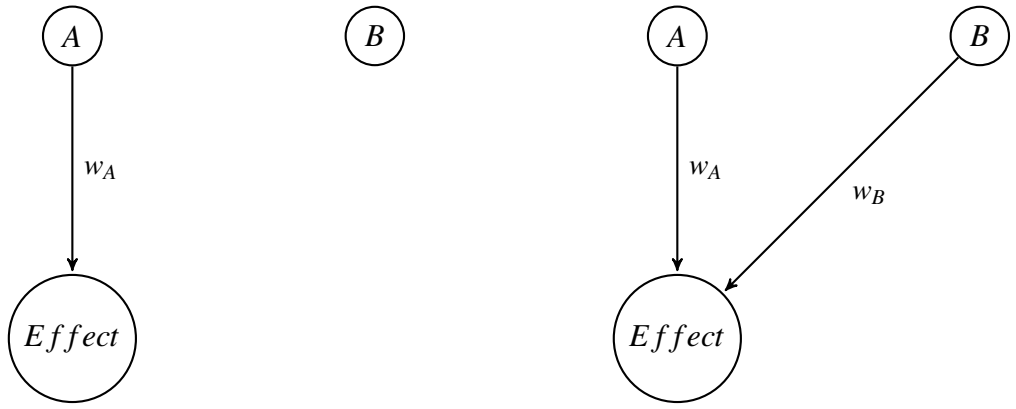


Fig. 6.1 Left: Single-cause graphical model. Right: Two-cause graphical model.

Support & Associative Weights

The conditional probability of an outcome occurring, given a graphical model G denoted $P(\text{outcome}|G)$, is equal to:

$$P(\text{outcome}|G) = \int_0^1 \cdots \int_0^1 P(\text{outcome}|w_1, w_2, \cdots, w_n, G) \cdot P(w_1, w_2, \cdots, w_n|G)\, \mathrm{d}w_1 \mathrm{d}w_2 \cdots \mathrm{d}w_n \tag{6.1}$$

Here w_n parameters are used to connect n nodes to the outcome in the graphical model. This integral can be approximated using Monte Carlo and the law of large numbers as:

$$P(\text{outcome}|G) \approx \frac{1}{m} \sum_{i=1}^{m} P(\text{outcome}|w_1^{(i)}, w_2^{(i)}, \cdots, w_n^{(i)}, G) \tag{6.2}$$

where $w_j^{(i)}$ is sampled from the distribution $P(w_1, w_2, \cdots, w_n|G)$.

We assume nodes of the graphical model correspond to stimuli representations, and parameters w_j correspond to associative weights from stimuli representations to an US representation, and are furthermore adjusted according to a competitive error-correction learning rule. If this is the case, then if $t \to \infty$ presentations of a stimuli-US contingency are presented such that the associative weights reach an approximate asymptote by trial a, then assuming that weights tend towards asymptotic values the probability of the asymptotic weights given the graphical model, $P(w_1^a, w_2^a, \cdots, w_n^a|G)$, will approach 1 as t approaches ∞. Thus as training progresses, the probability of sampling $\left\{w_1^a, w_2^a, \cdots, w_n^a\right\}$ from the distribution $P(w_1, w_2, \cdots, w_n|G)$ approaches 1. Therefore if we take the Monte Carlo approximation to be sampling from the weight values over conditioning trials, and the trials going on infinitely, then the Monte Carlo approximation approaches:

$$P(\text{outcome}|G) \approx \frac{1}{T} \sum_{i=1}^{T} P(\text{outcome}|w_1^a, w_2^a, \cdots, w_n^a, G) \tag{6.3}$$

for $T \to \infty$.

This is approximate to calculating 1 minus the average error in predicting the occurrence of the outcome given associative weights with asymptotic values:

$$P(\text{outcome}|G) \approx (1 - \bar{\delta}_{\text{outcome}}|w_1^a, w_2^a, \cdots, w_n^a, G) \tag{6.4}$$

Therefore the support value between two graphs G_1 and G_2 becomes:

$$log(\frac{P(\text{outcome}|G_1)}{P(\text{outcome}|G_2)}) \approx log\frac{(1 - \bar{\delta}_{\text{outcome}}|G_1)}{(1 - \bar{\delta}_{\text{outcome}}|G_2)} \tag{6.5}$$

and since:

$$log(a/b) = log(a) - log(b) \tag{6.6}$$

we have:

$$log(\frac{P(\text{outcome}|G_1)}{P(\text{outcome}|G_2)}) \approx log(1 - \bar{\delta}_{\text{outcome}}|G_1) - log(1 - \bar{\delta}_{\text{outcome}}|G_2) \tag{6.7}$$

If G_1 is a sub-graph of G_2, that is G_1 is a subset of the stimuli and their weights contained in G_2, then this expression captures the proportion of the US occurrence which G_1 predicts compared to all stimuli representations. This is equivalent to the combined weights of all stimuli representations in G_1. Thus Bayesian support bears a relationship to weights between stimuli representations.

This connection of associative learning to Bayesian graphical models is of interest as the latter have been tied theoretically to, for instance, the connectome of the human brain through imaging data (Bullmore and Bassett, 2011). Thus, models of learning that instantiate

principles of Bayesian graphical models by calculating causal dependency between all cues (not just reinforcers) through associations, such as the DE model, could describe how the brain instantiates approximate Bayesian inference through associations learning. Thus, both the general connectionist framework of the model as well as its learning rule, which leads to revaluation of past associations in a form similar to Bayes' rule, instantiate processes of Bayesian inference.

Recurrent Network DE

Finally, two experiments in instantiating components of the DE model as recurrent neural networks have been experimented upon, detailed in the next two subsections.

Having demonstrated the ability of the DE model in predicting numerous phenomena in the field of learning theory, we sought as an avenue of future work to demonstrate that its principles could be extended to produce a model of learning with applicability to machine learning problems. Thus, we instantiated the model using principles of predictive-coding. Predictive coding is the principle, derived from cortical modelling, that a neural network attempts to minimize the free-energy or input-error of a given layer, and any residual energy/error is passed on for the next layer to account for (Friston and Kiebel, 2009; Rao and Ballard, 2004). Hence, the activity of a given node encodes both its input activation from the layers below it, and a predictive component from other nodes. Thus its activity in essence encodes an error-term for that node. A simplified representation of such an architecture is presented in Figure 6.2.

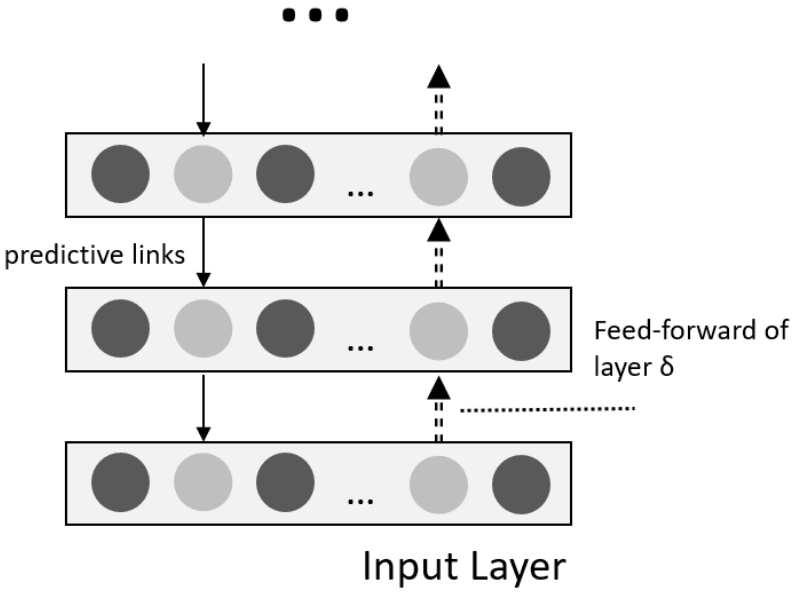


Fig. 6.2 Predictive coding architecture.

The advantage of predictive coding is that it avoids many of the pitfalls inherent to backpropagation. For instance, since the network attempts to predict all its inputs and hence learn a robust representation of the interdependencies in the input, it is more suited for unsupervised learning. As we were interested in maintaining neurobiological plausibility in the operation of this deep network, we further agreed with claims that there are several difficulties in claiming that backpropagation is instantiated in the brain. For instance, it is unclear how learning taking place between sets of neurons can accommodate the backpropagation of error-signals over multiple neurons, and how discrete neuronal spikes can be differentiated, though progress is being made in this area (Richards, 2017). Further, the predictive coding network demonstrated in this section uses deep layers of hidden units, the ReLU activation function, and real-time inputs (that is, the activity of the network is completely determined by its history of inputs).

The end result is a deep recurrent neural network extension of the DE model, which we have applied to a natural language learning task. It incorporates rules for learning weights from a child to a parent in a Hebbian-like network that incorporates an element of competition between nodes (through both the error-correction rule utilized, and the rule for learning child to parent links). Competition between nodes has been argued to be necessary in a Hebbian network to avoid run-away growth of connections (Zenke et al., 2017). Thus, the learning rules operate in such manner as to produce sparse representations.

$$\mathscr{A}_i^t = |Y_i^t - \hat{Y}_i^t \cdot \vartheta| \tag{6.8}$$

where A is the activity of node i at time t, Y_i^t is the input activity of node i, $\hat{Y}_i^t$ is the prediction produced for this node by other nodes, and ϑ is an associative retrieval discount

parameter. As the absolute value is taken, this ensures no negative activities can occur. This in principle means that the activation function of the network is a variation of the Rectifier Linear Unit (ReLU). The ReLU activation function has found utility in machine learning due to its straightforward derivative simplifying backpropagation (though in this experiment no backpropagation was used). It is computationally cheap to calculate, and tends to produce sparse representations. It has been analyzed in the context of neuroscience in its relation to the stability of synaptic dynamics (Hahnloser and Seung, 2001).

The input or associative activation to a given node in the deep network are both calculated using the following equation:

$$Y_i^t / \hat{Y}_i^t = \sum_j w_{j \to i}^t \cdot \mathscr{A}_j^t \tag{6.9}$$

Here w is the weight, $Y / \hat{Y}$ is the input/associative activation, and $\mathscr{A}$ is the activity. These predictions are made respectively over all nodes in the layer below, or of the same layer and the layer above it (if such a layer exists).

For the input layer, inputs are read directly as character values from the .txt file in order of appearance. The .txt file used was the Adventures of Sherlock Holmes, available publicly from https://www.gutenberg.org/ebooks/1661. All alphabetical characters and space/new-line characters are converted to lower-case characters. All characters besides these are omitted from the inputs.

The predictive links of the model are calculated using the following weight-update equation:

$$w_{i\to j}^{t+1} = w_{i\to j}^{t} + \alpha \mathscr{A}_i^t (\lambda_{i\to j}^t - \sum_l w_{l\to j}^t \cdot \mathscr{A}_l^t) \tag{6.10}$$

Here α is the learning rate, $\mathscr{A}$ is the predictor activity, and λ is the dynamic asymptote of learning.

The asymptote of learning is calculated using the dynamic asymptote rule:

$$\lambda_{i\to j}^t = \mathscr{A}_j^t - |\mathscr{A}_j^t - \mathscr{A}_i^t| \tag{6.11}$$

The top-up hidden-layer input links (feed-forward links) are calculated using a rule, which promotes sparse and non-redundant activation of hidden layer nodes to promote robust representations capable of aiding in the learning of temporal correlations between inputs:

$$w_{i\to j}^{t+1} = w_{i\to j}^{t} + \alpha \mathscr{A}_i^t (\omega_i^t + noise) \tag{6.12}$$

This omega is calculated using a similarity measure as follows:

$$\omega_i^t = min(\mathscr{A}_i^t - avg.\ layer\ act., F_i^t - avg.\ F.) \tag{6.13}$$

Hence a node will receive a stronger input from the layer below it if it has above average activity when compared to the average node in its layer, and when its F value is higher than the average in its layer. Here F_i denotes the functional similarity of node i towards the inputs of the layer below it. What this variable measures is how well i predicts the current configuration of inputs in the layer below it, with this functional similarity being weighted by the unexpectedness of the input nodes. There is a two-fold motivation for this weight-update

rule: firstly it promotes sparsity of activity in hidden layers, which conforms to the idea that a given event should only have few principle causes, and that a given episode of learning can be broken into principle components. The second is that hidden-layer representations should aid in predicting the layer below themselves. Hence nodes which achieve this more effectively receive stronger inputs (which due to the weight update are also weighted by the activity of the nodes from which they receive feed-forward inputs), making them more likely to be active in similar situations in the future.

$$F_i^t = \frac{1}{\#j}\sum_j (sign(w_{i\to j}^t) == sign(\delta_j^t)) \cdot |(w_{i\to j}^t)| \cdot |\delta_j^t| \tag{6.14}$$

The overall architecture used for the recurrent neural network is visible in Figure 6.3. The network is composed of one input layer, and four hidden layers. Each hidden layer consists of 540 units. The input layer has 27 units, each of which encodes an alphabetic character or the newline/space characters. Units in each layer predict each other, the layer below them, and the input layer. Hence the input layer receives predictive links from all hidden layers. The predictive links of the model are updates according to the traditional delta/double-error rule, while the feed-forward links changed according to the aforementioned sparsity-maintaining weight-update rule.

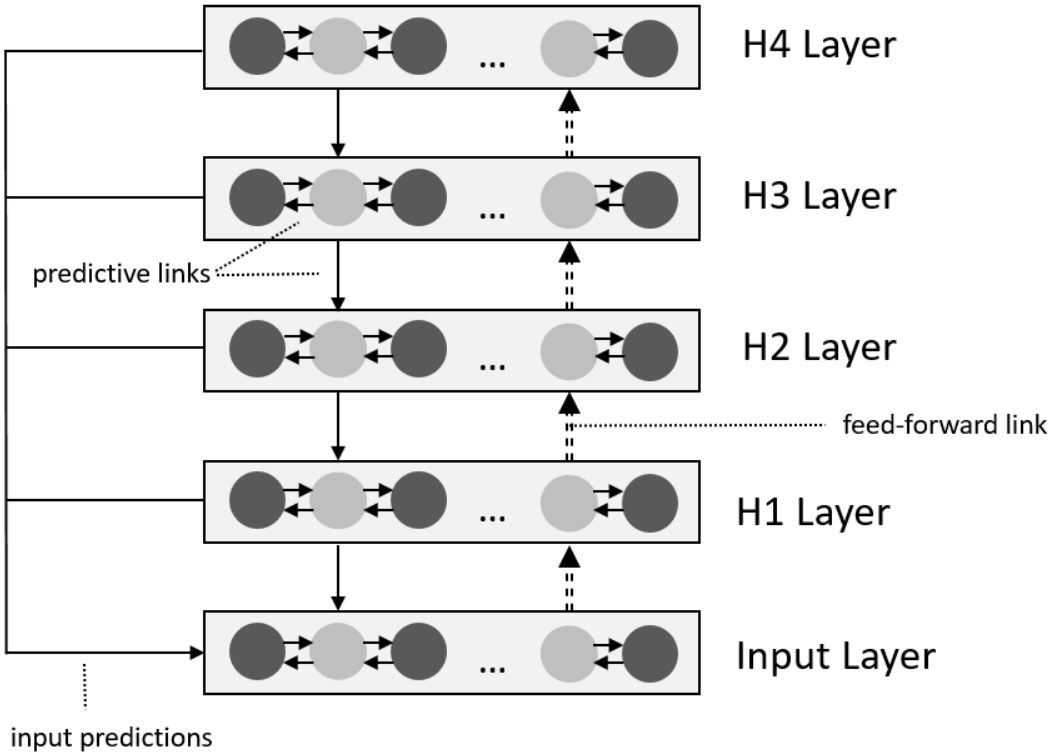


Fig. 6.3 The deep recurrent network architecture used on the natural language task.

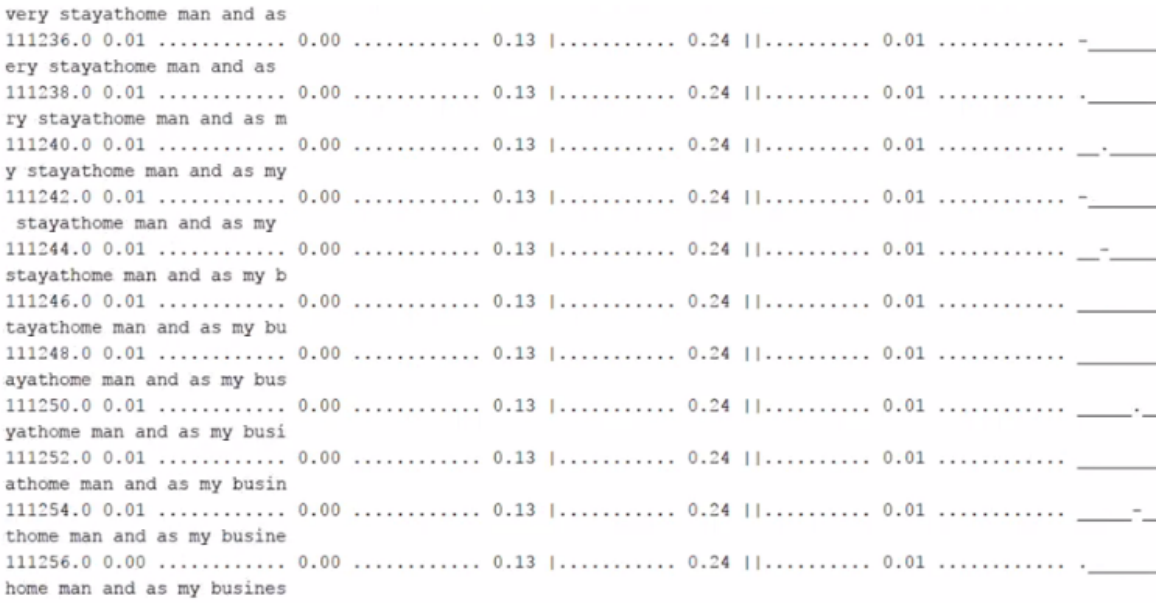


Fig. 6.4 Sample predictions of the deep double-error network on Sherlock Holmes text file (available publicly from https://www.gutenberg.org/ebooks/1661). The lack of a 'backwards-in-time' discount for learning between a predictor that became active after an outcome produces degenerate learning wherein inputs learn to predict themselves with a small delay.

One pitfall in this predictive framework is the importance of encoding time-dependence. This was done through a discount of the asymptote of learning from the predictor to the outcome in cases wherein the outcome was present before the predictor. Without such a discount, the network quickly learned to predict the inputs by learning a hidden-unit representation such that in effect inputs predicted themselves through the hidden layer. The result of this was that the network predicted its own input 1-to-1 with a two step delay (Figure 6.4).

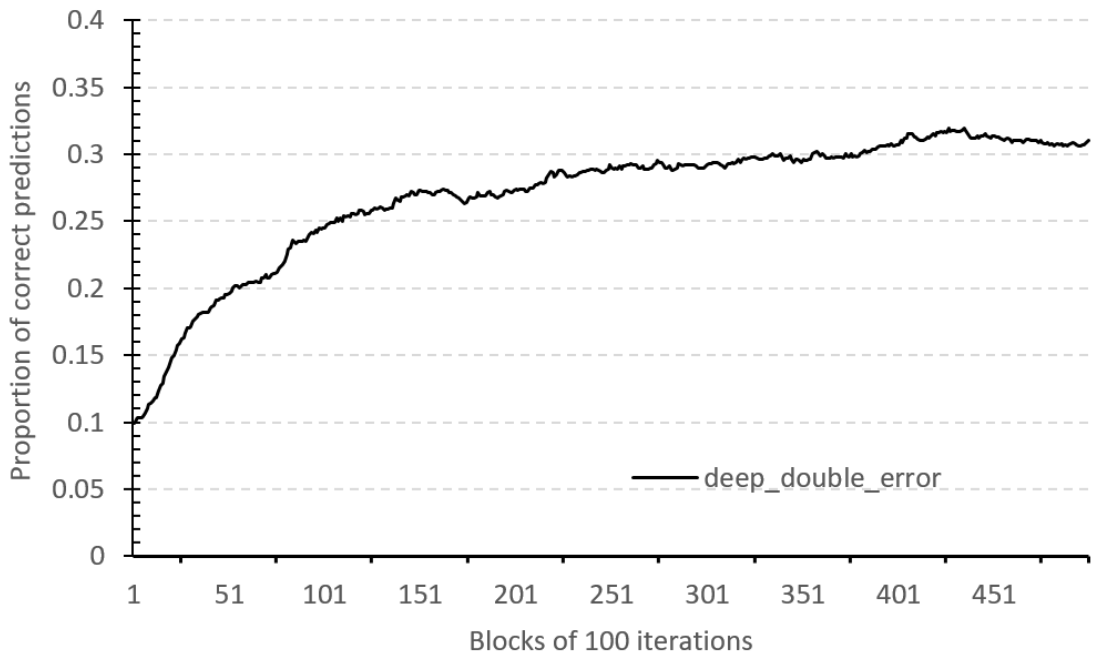


Fig. 6.5 Proportion of correct prediction of the deep DE model on Sherlock Holmes text file.

However once rectified, the deep DE model managed to perform above random chance, achieving a proportion of correct character predictions of 30% after 50 thousand iterations (Figure 6.5). This demonstrates that the unique core components of the DE model, namely its predictor error-term and dynamic asymptote, can be extended to a deep recurrent neural network, which is capable of predicting highly complex sequences of alphabetic characters with a degree of success.

Adagrad RNN NLP Experiment

In the previous section of proof-of-concept future work we introduced a deep, recurrent network implementation of the DE model based on principles of predictive coding. In this section, as we sought to demonstrate that the principles of the DE model are also applicable to more standard recurrent networks, I have adapted an open-source python RNN to include modulation by a predictor error-term. Though both the previous extension of the DE model and the one introduced in this section perform a similar character-by-character prediction of a text file, the performance measures used are not directly comparable due to the latter using cross entropy loss, while the former uses a normalized MSE. The architecture of the vanilla RNN used in this section is depicted in Figure 6.6.

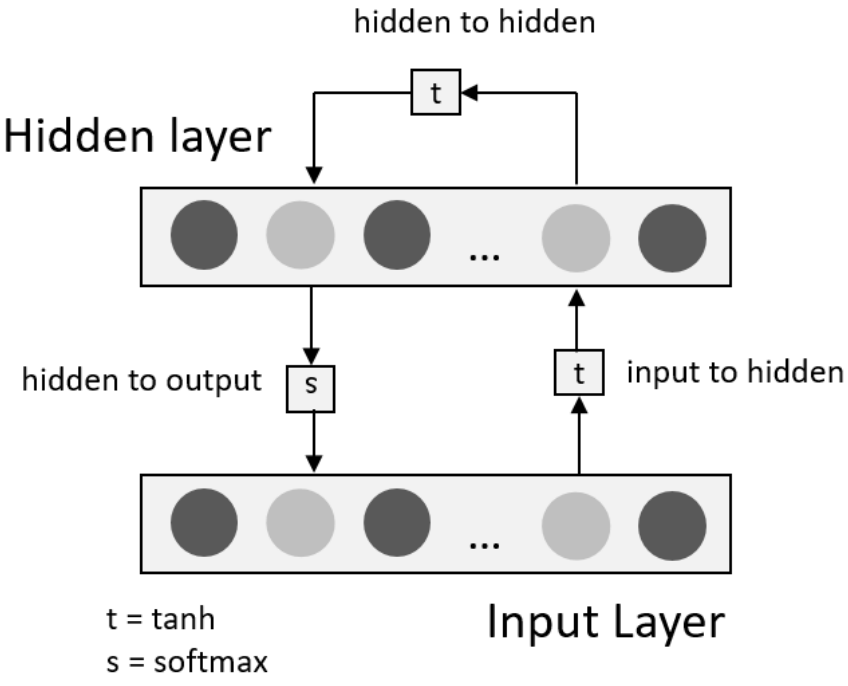


Fig. 6.6 Architecture of the open-source RNN used.

The RNN is written by Andrej Karpathy (Karpathy, 2017), and learns text files character by character. For its loss function, the network uses the cross-entropy loss of the characters of the file:

$$loss = \sum_i -log(p(i)) \tag{6.15}$$

where $p(i)$ is the predicted probability of the next character to appear in the sampled section of the text. These probabilities are calculated as the dot product of hidden to output weights with the hidden activities, which are passed through the softmax function:

$$\mathbf{p} = e^{\frac{\mathbf{Why} \cdot \mathbf{h}}{\sum \mathbf{Why} \cdot \mathbf{h}}} \tag{6.16}$$

The activation function of the hidden units of the network uses the hyperbolic tangent function:

$$\mathbf{h} = tanh(\mathbf{Wxh} \cdot \mathbf{x} + \mathbf{Whh} \cdot \mathbf{h} + bh) \tag{6.17}$$

where **h** are the hidden-state activations, Wxh are the input to hidden weights, Whh are the hidden to hidden weights, bh is the hidden bias, and $\cdot$ indicates that the cross product is taken.

For the weight update, the RNN utilizes the adagrad extension of stochastic gradient descent, which generally speaking increases the rate of learning for more sparse parameters.

We modified this base architecture by defining a set of hidden to hidden associative weights, which are calculated using the Double-Error error-correction rule:

$$\Delta Wha_{i \to j} = \sum_{t} \alpha(h_j - |h(j) - h(i)| - p(j))|(h_i - p(i)) \tag{6.18}$$

The predictor-error terms calculated using these associative weights are subsequently used to modulate the Δ**Why** hidden to output changes in the backpropagation step of the algorithm:

$$\Delta \mathbf{Why} = \mathbf{dy} \cdot (\mathbf{h} \times \delta_{\mathbf{h}}).T \tag{6.19}$$

where **dy** is the backpropagated targets, $\delta_{\mathbf{h}}$ is the predictor error-terms, T is the transpose, and $\times$ denotes element-wise multiplication.

When compared to the vanilla implementation of the RNN, the modified algorithm performs slightly better after 100x100 iterations over the text (Figure 6.7).

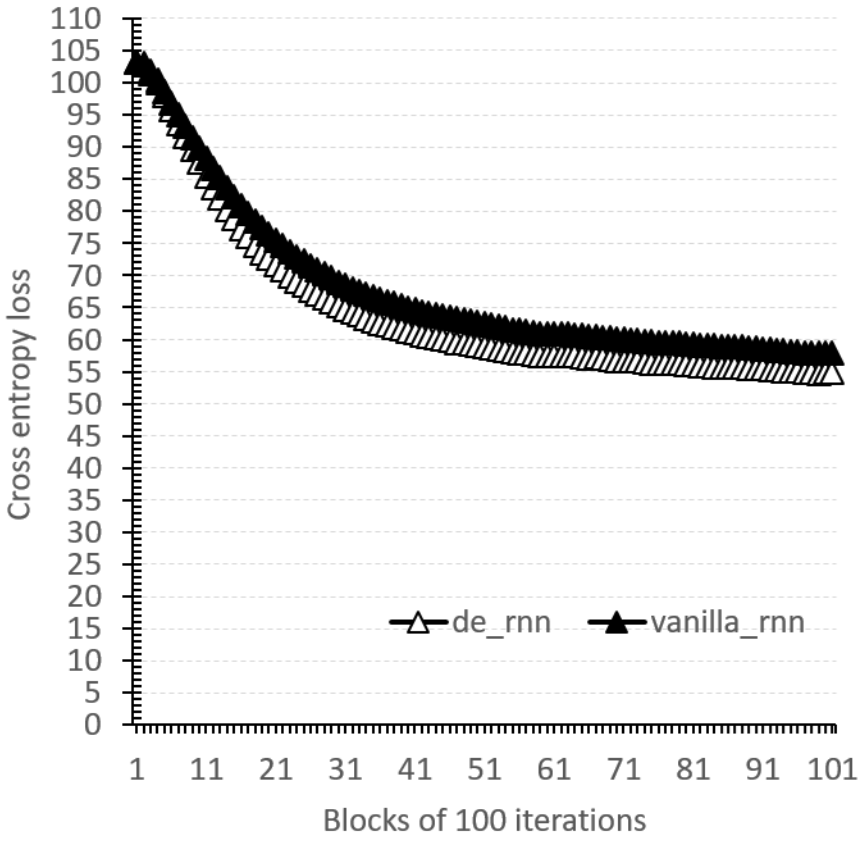


Fig. 6.7 Performance (cross entropy loss) of the vanilla RNN vs. the same RNN with hidden layer predictor error-terms modulating the backpropagation of errors through the hidden layer.

References

Ahveninen, J., Kähkönen, S., Tiitinen, H., Pekkonen, E., Huttunen, J., Kaakkola, S., Ilmoniemi, R. J., and Jääskeläinen, I. P. (2000). Suppression of transient 40-Hz auditory response by haloperidol suggests modulation of human selective attention by dopamine D2 receptors. *Neuroscience Letters*, 292(1):29–32.

Alonso, E., Sahota, P., and Mondragon, E. (2014). Computational Models of Classical Conditioning - A Qualitative Evaluation and Comparison. In Duval, B., van den Herik, J., Loiseau, S., and Filipe, J., editors, *Proceedings of the 6th International Conference on Agents and Artificial Intelligence*, pages 544–547, Setúbal, Portugal. SCITEPRESS - Science and and Technology Publications.

Alonso, E. and Schmajuk, N. (2012). Special issue on computational models of classical conditioning guest editors' introduction. *Learning & Behavior*, 40(3):231–240.

Amundson, J. C. and Miller, R. R. (2008). CS-US temporal relations in blocking. *Learning & Behavior*, 36(2):92–103.

Arcediano, F., Matute, H., and Miller, R. R. (1997). Blocking of Pavlovian conditioning in humans. *Learning and Motivation*, 28(2):188–199.

Arcediano, F. and Miller, R. R. (2002). Some constraints for models of timing: A temporal coding hypothesis perspective. *Learning and Motivation*, 33(1):105–123.

Baker, A. G., Murphy, R. A., and Mehta, R. (2003). Learned irrelevance and retrospective correlation learning. *The Quarterly Journal of Experimental Psychology: Section B*, 56(1):90–101.

Balkenius, C. and Morén, J. (1998). Computational models of classical conditioning: A comparative study. In Pfeifer, R., Blumberg, B., Meyer, J.-A., and Wilson, S. W., editors, *Proceedings of the fifth international conference on simulation of adaptive behavior on From animals to animats*, Vol. 5, pages 348–353. MIT Press, Cambridge, MA, USA.

Barria, A., Muller, D., Derkach, V., Griffith, L. C., and Soderling, T. R. (1997). Regulatory phosphorylation of AMPA-type glutamate receptors by CaM-KII during long-term potentiation. *Science*, 276(5321):2042–2045.

Blair, C. A. J. and Hall, G. (2003). Perceptual learning in flavor aversion: Evidence for learned changes in stimulus effectiveness. *Journal of Experimental Psychology: Animal Behavior Processes*, 29(1):39.

Blaisdell, A. P., Denniston, J. C., and Miller, R. R. (1998). Temporal encoding as a determinant of overshadowing. *Journal of Experimental Psychology: Animal Behavior Processes*, 24(1):72.

Blough, D. S. (1975). Steady state data and a quantitative model of operant generalization and discrimination. *Journal of Experimental Psychology: Animal Behavior Processes*, 1(1):3–21.

Bonardi, C. and Hall, G. (1996). Learned irrelevance: No more than the sum of CS and US preexposure effects? *Journal of Experimental Psychology: Animal Behavior Processes*, 22(2):183.

Bradfield, L. and McNally, G. P. (2008). Unblocking in Pavlovian fear conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, 34(2):256.

Brandon, S. E., Vogel, E. H., and Wagner, A. R. (2000). A componential view of configural cues in generalization and discrimination in Pavlovian conditioning. *Behavioural Brain Research*, 110(1):67–72.

Brandon, S. E., Vogel, E. H., and Wagner, A. R. (2003). Stimulus representation in SOP: I. Theoretical rationalization and some implications. *Behavioural Processes*, 62(1-3):5–25.

Brogden, W. J. (1939). Sensory pre-conditioning. *Journal of Experimental Psychology*, 25(4):323–332.

Bullmore, E. T. and Bassett, D. S. (2011). Brain graphs: graphical models of the human brain connectome. *Annual Review of Clinical Psychology*, 7:113–140.

Bush, R. R. and Mosteller, F. (1955). *Stochastic models for learning.* Wiley, New York, USA.

Caruana, R. (1998). A dozen tricks with multitask learning. In Orr, G. and Müller, K. R., editors, *Neural Networks: Tricks of the Trade*, chapter 8, pages 165–191. Springer-Verlag, Berlin Heidelberg, Germany.

Channell, S. and Hall, G. (1983). Contextual effects in latent inhibition with an appetitive conditioning procedure. *Learning & Behavior*, 11(1):67–74.

Colwill, R. M. and Motzkin, D. K. (1994). Encoding of the unconditioned stimulus in Pavlovian conditioning. *Learning & Behavior*, 22(4):384–394.

Corlett, P. R., Murray, G. K., Honey, G. D., and Aitken, M. R. F. (2007). Disrupted prediction-error signal in psychosis: evidence for an associative account of delusions. *Brain*, 130(9):2387–2400.

Courrieu, P. (2008). Solving time of least square systems in Sigma-Pi unit networks. *arXiv preprint arXiv:0804.4808*.

Cueto, M. A., Morton, J., and Sturmfels, B. (2010). Geometry of the restricted Boltzmann machine. In Viana, M. A. G. and Wynn, H. P., editors, *Algebraic Methods in Statistics and Probability*, volume 516, chapter 11, pages 135–153. AMS, Contemporary Mathematics, Champaign, IL, USA.

Danks, D. (2003). Equilibria of the Rescorla–Wagner model. *Journal of Mathematical Psychology*, 47(2):109–121.

Delamater, A. R. and Holland, P. C. (2008). The influence of CS-US interval on several different indices of learning in appetitive conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, 34(2):202.

Dickinson, A. (1977). Appetitive–aversive interactions: superconditioning of fear by an appetitive CS. *The Quarterly Journal of Experimental Psychology*, 29(1):71–83.

Dickinson, A. (1996). Within compound Associations Mediate the Retrospective Revaluation of Causality Judgements. *The Quarterly Journal of Experimental Psychology Section B*, 49(1):60–80.

Dickinson, A., Hall, G., and Mackintosh, N. J. (1976). Surprise and the attenuation of blocking. *Journal of Experimental Psychology: Animal Behavior Processes*, 2(4):313–322.

Donegan, N. H. and Wagner, A. R. (1987). Conditioned diminution and facilitation of the UR: A sometimes opponent-process interpretation. In Gormezano, I., Prokasy, W. F., and Thompson, R. F., editors, *Classical Conditioning*, chapter 13, pages 339–369. Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, USA.

Dwyer, D. M. (1999). Retrospective revaluation or mediated conditioning? The effect of different reinforcers. *The Quarterly Journal of Experimental Psychology: B, Comparative and physiological psychology*, 52(4):289–306.

Espinet, A., Iraola, J. A., Bennett, C. H., and Mackintosh, N. J. (1995). Inhibitory associations between neutral stimuli in flavor-aversion conditioning. *Animal Learning & Behavior*, 23(4):361–368.

French, R. M. (1992). Semi-distributed representations and catastrophic forgetting in connectionist networks. *Connection Science*, 4(3-4):365–377.

Friston, K. and Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 364(1521):1211–1221.

Gallistel, C. R., Fairhurst, S., and Balsam, P. (2004). The learning curve: implications of a quantitative analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 101(36):13124–31.

Gallistel, C. R. and Gibbon, J. (2000). Time, rate, and conditioning. *Psychological Review*, 107(2):289–344.

Gheorghescu, A., Mondragón, E. and Alonso, E. (2017). Pearce Model Simulator (Version 1)[Computer Software]. St. Albans, UK: CAL-R. Available from https://www.cal-r.org/index.php?id=Pearce-sim.

Ghirlanda, S. (2015). On elemental and configural models of associative learning. *Journal of Mathematical Psychology*, 64:8–16.

Ghirlanda, S. and Ibadullayev, I. (2015). Solution of the comparator theory of associative learning. *Psychological Review*, 122(2):242.

Ghorashi, A., Mondragón, E. and Alonso, E. (2017). Replaced Elements Model Simulator (Version 1)[Computer Software]. St. Albans, UK: CAL-R. Available from https://www.cal-r.org/index.php?id=REM-sim

Ginsburg, S. and Jablonka, E. (2010). The evolution of associative learning: A factor in the Cambrian explosion. *Journal of Theoretical Biology*, 266(1):11–20.

Glautier, S. (2013). Revisiting the learning curve (once again). *Frontiers in Psychology*, 4:982.

Glautier, S., Redhead, E., Thorwart, A., and Lachnit, H. (2010). Reduced Summation with Common Features in Causal Judgments. *Experimental Psychology*, 57(4):252–259.

Gomez, M., De Castro, E., Guarin, E., Sasakura, H., Kuhara, A., Mori, I., Bartfai, T., Bargmann, C. I., and Nef, P. (2001). Ca2+ Signaling via the Neuronal Calcium Sensor-1 Regulates Associative Learning and Memory in C. elegans. *Neuron*, 30(1):241–248.

Gray, J., Alonso, E., Mondragón, E., and Fernández, A. (2012). Temporal Difference Simulator(Version 1)[Computer Software]. St. Albans, UK: CAL-R. Available from https://www.cal-r.org/index.php?id=TD-sim

Griffiths, T. L., Kemp, C., and Tenenbaum, J. B. (2008). Bayesian Models of Cognition. In Sun, R., editor, *The Cambridge Handbook of Computational Psychology*, chapter 3, pages 59–100. Cambridge University Press, UK.

Grikietis, R., Mondragón, E., and Alonso, E. (2016). Pearce & Hall Simulator (Version 1)[Computer Software]. St. Albans, UK: CAL-R. Available from https://www.cal-r.org/index.php?id=PHsim

Hahnloser, R. H. R. and Seung, H. S. (2000). Permitted and forbidden sets in symmetric threshold-linear networks. In *Advances in Neural Information Processing Systems 13*, Eds. Dietterich, T.G., Becker, S., and Ghahramani, Z., pages 217–223. MIT Press, Cambridge, MA, USA.

Hall, G. (1991). *Perceptual and Associative Learning.* Clarendon Press/Oxford University Press, UK.

Hall, G. and Honey, R. C. (1989). Contextual effects in conditioning, latent inhibition, and habituation: associative and retrieval functions of contextual cues. *Journal of Experimental Psychology: Animal Behavior Processes*, 15(3):232.

Hall, G. and Pearce, J. M. (1979). Latent inhibition of a CS during CS-US pairings. *Journal of Experimental Psychology. Animal Behavior Processes*, 5(1):31–42.

Hall, G. and Rodriguez, G. (2010). Associative and nonassociative processes in latent inhibition: An elaboration of the Pearce-Hall model. In Lubow, R. E. and Weiner, I., editors, *Latent inhibition: Cognition, Neuroscience and Applications to Schizophrenia*, chapter 6, pages 114–136. Cambridge University Press, New York, USA.

Harris, J. A. (2006). Elemental Representations of Stimuli in Associative Learning. *Psychological Review*, 113(3):584–605.

Harris, J. A. and Livesey, E. J. (2010). An attention-modulated associative network. *Learning & Behavior*, 38(1):1–26.

Hassabis, D., Kumaran, D., Summerfield, C., and Botvinick, M. (2017). Neuroscience-inspired artificial intelligence. *Neuron*, 95(2):245–258.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8):1735–1780.

Holland, P. C. (1983). Representation-mediated overshadowing and potentiation of conditioned aversions. *Journal of Experimental Psychology: Animal Behavior Processes*, 9(1):1–13.

Holland, P. C. and Forbes, D. T. (1982). Representation-mediated extinction of conditioned flavor aversions. *Learning and Motivation*, 13(4):454–471.

Holland, P. C. and Schiffino, F. L. (2016). Mini-review: Prediction errors, attention and associative learning. *Neurobiology of Learning and Memory*, 131:207–215.

Hull, L. C. (1943). *Principles of behavior: an introduction to behavior theory.* Appleton-Century, 1 edition. New York, USA.

Jennings, D. and Kirkpatrick, K. (2006). Interval duration effects on blocking in appetitive conditioning. *Behavioural Processes*, 71:318–329.

Jie, H. L. (2008). Neuroimaging of associative learning. *Masters Thesis*. National University of Singapore, Singapore.

Kamin, L. J. (1968). "Attention-like" processes in classical conditioning. In Jones, M. R., editor, *Miami symposium on the prediction of behavior: Aversive stimulation*, pages 9–31, Coral Cables, Florida. University of Miami Press, USA.

Kamin, L. J. (1969). Selective association and conditioning. *Fundamental Issues in Associative Learning*, pages 42–64. Dalhousie University, Halifa, Canada.

Karpathy, A. (2017). min-char-rnn.py. (Version 1)[Computer Software] Available from https://gist.github.com/karpathy/d4dee566867f8291f086

Kehoe, E. J., Schreurs, B. G., and Graham, P. (1987). Temporal primacy overrides prior training in serial compound conditioning of the rabbit's nictitating membrane response. *Animal Learning & Behavior*, 15(4):455–464.

Kobayashi, K. and Poo, M. (2004). Spike Train Timing-Dependent Associative Modification of Hippocampal CA3 Recurrent Synapses by Mossy Fibers. *Neuron*, 41(3):445–454.

Koblitz, N. (1996). p-adic Numbers, p-adic Analysis, and Zeta-Functions. page 172. Springer. New York, USA.

Kohler, E. A. and Ayres, J. J. B. (1979). The Kamin blocking effect with variable-duration CSs. *Animal Learning & Behavior*, 7(3):347–350.

Kruschke, J. K. (2009). Models of Attentional Learning. In Pothos, E. M. and Wills, A. J., editors, *Formal Approaches in Categorization*, pages 120–152. Cambridge University Press, Cambridge, UK.

Kruse, J. M., Overmier, J. B., Konz, W. A., and Rokke, E. (1983). Pavlovian conditioned stimulus effects upon instrumental choice behavior are reinforcer specific. *Learning and Motivation*, 14(2):165–181.

Kutlu, M. G. and Schmajuk, N. A. (2012). Solving Pavlov's puzzle: Attentional, associative, and flexible configural mechanisms in classical conditioning. *Learning & Behavior*, 40(3):269–291.

Lachnit, H., Schultheis, H., König, S., Üngör, M., and Melchers, K. (2008). Comparing elemental and configural associative theories in human causal learning: A case for attention. *Journal of Experimental Psychology: Animal Behavior Processes*, 34(2):303–313.

Lattal, K. M. and Nakajima, S. (1998). Overexpectation in appetitive Pavlovian and instrumental conditioning. *Learning & behavior*, 26(3):351–360.

Le Pelley, M. E. (2004). The role of associative history in models of associative learning: A selective review and a hybrid model. *The Quarterly Journal Of Experimental Psychology*, 57B(3):193–243.

Le Pelley, M. E. and McLaren, I. P. L. (2001). Retrospective revaluation in humans: Learning or memory? *The Quarterly Journal of Experimental Psychology: Section B*, 54(4):311–352.

LeCun, Y. A., Bottou, L., Müller, K. R., and Orr, G. B. (2002). Efficient backprop. In Montavon, G., Orr, G. B., and Müller, K. R., editors, *Neural Networks: Tricks of the Trade*, chapter 1, pages 9–50. Springer-Verlag, Berlin Heidelberg, Germany.

Leung, H. T., Killcross, A. S., and Westbrook, R. F. (2011). Additional exposures to a compound of two preexposed stimuli deepen latent inhibition. *Journal of Experimental Psychology: Animal Behavior Processes*, 37(4):394.

Lober, K. and Lachnit, H. (2002). Configural learning in human Pavlovian conditioning: acquisition of a biconditional discrimination. *Biological Psychology*, 59(2):163–168.

Lu, W. Y., Man, H. Y., Ju, W., Trimble, W. S., MacDonald, J. F., and Wang, Y. T. (2001). Activation of synaptic NMDA receptors induces membrane insertion of new AMPA receptors and LTP in cultured hippocampal neurons. *Neuron*, 29(1):243–254.

Lubow, R. E. (1965). Effects of frequency of nonreinforced preexposure of the cs. *Journal of Comparative and Physiologial Psychology*, 60(3):454–457.

Ludvig, E. A., Bellemare, M. G., and Pearson, K. G. (2011). A primer on reinforcement learning in the brain: Psychological, computational, and neural perspectives. In Alonso, E. and Mondragon, E., editors, *Computational Neuroscience for Advancing Artificial Intelligence: Models, Methods and Applications*, chapter 6, pages 111–144. IGI Global.

Ludvig, E. A. and Koop, A. (2008). Learning to Generalize through Predictive Representations: A Computational Model of Mediated Conditioning. In *From Animals to Animats 10*, pages 342–351, Berlin, Heidelberg. Springer Berlin Heidelberg, Germany.

Ludvig, E. A., Mirian, M. S., Kehoe, E. J., and Sutton, R. S. (2017). Associative Learning from Replayed Experience. *bioRxiv preprint*.

Ludvig, E. A., Sutton, R. S., Verbeek, E., and Kehoe, E. J. (2009). A computational model of hippocampal function in trace conditioning. In Koller, D., Schuurmans, D., Bengio, Y., and Bottou, L., editors, *Advances in Neural Information Processing Systems 21*, pages 993–1000. MIT Press, Cambridge, MA, USA.

Maass, W. (2010). Liquid state machines: motivation, theory, and applications. In Cooper, S. B. and Sorbi, A., editors, *Computability in context: computation and logic in the real world*, chapter 8, pages 275–296. World Scientific Publishing Company, London, UK.

Mackintosh, N. J. (1973). Stimulus selection: Learning to ignore stimuli that predict no change in reinforcement. In Hinde, R. A. and Stevenson-Hinde, J., editors, *Constraints on Learning: Limitations and Predispositions*. Academic Press, Oxford, UK.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, 82(4):276–298.

Mackintosh, N. J. (1976). Overshadowing and stimulus intensity. *Animal Learning & Behavior*, 4(2):186–192.

Mackintosh, N. J., Kaye, H., and Bennett, C. H. (1991). Perceptual learning in flavour aversion conditioning. *The Quarterly Journal of Experimental Psychology*, 43(3):297–322.

McLaren, I. P. L. (1993). APECS: A solution to the sequential learning problem. In *Proceedings of the xvth Annual Convention of the Cognitive Science Society*, pages 717–722. University of Colorado in Boulder, Colorado, USA.

McLaren, I. P. L. and Mackintosh, N. J. (2000). An elemental model of associative learning: I. Latent inhibition and perceptual learning. *Animal Learning & Behavior*, 28(3):211–246.

Miller, R. and Matzel, L. D. (1988). The Comparator Hypothesis: A Response Rule for The Expression of Associations. *Psychology of Learning and Motivation*, 22:51–92.

Miller, R. R., Barnet, R. C., and Grahame, N. J. (1995). Assessment of the Rescorla-Wagner model. *Psychological Bulletin*, 117(3):363–86.

Miller, R. R. and Witnauer, J. E. (2016). Retrospective revaluation: The phenomenon and its theoretical implications. *Behavioural Processes*, 123:15–25.

Mitchell, C. J., De Houwer, J., and Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences*, 32(2):183–198.

Mondragón, E., Alonso, E., Fernández, A., and Gray, J. (2013a). An extension of the Rescorla and Wagner Simulator for context conditioning. *Computer Methods and Programs in Biomedicine*, 110(2):226–230.

Mondragón, E., Alonso, E., and Kokkola, N. (2017). Associative Learning Should Go Deep. *Trends in Cognitive Sciences*, 21(11):822–825.

Mondragón, E., Alonso, E., Leclerq, G., and Rosas, J. M. (1993). Discriminación y retención a corto plazo de la información temporal. In *V Reunión de la Sociedad Española de Psicología Comparada*, Barcelona, Spain.

Mondragón, E., Gray, J., and Alonso, E. (2013b). A Complete Serial Compound Temporal Difference Simulator for Compound stimuli, Configural cues and Context representation. *Neuroinformatics*, 11(2):259–261.

Mondragón, E., Gray, J., Alonso, E., Bonardi, C., and Jennings, D. J. (2014). SSCC TD: A Serial and Simultaneous Configural-Cue Compound Stimuli Representation for Temporal Difference Learning. *PLoS ONE*, 9(7):e102469.

Mondragón, E. and Hall, G. (2002). Analysis of the perceptual learning effect in flavour aversion learning: Evidence for stimulus differentiation. *The Quarterly Journal of Experimental Psychology: Section B*, 55(2):153–169.

Mondragón, E. and Murphy, R. A. (2010). Perceptual learning in an appetitive Pavlovian procedure: Analysis of the effectiveness of the common element. *Behavioural Processes*, 83(3):247–256.

Mondragón, E., Alonso, E., Fernández, A., and Gray, J. (2012). Rescorla & Wagner Model Simulator (Version 4)[Computer Software]. St. Albans, UK: CAL-R. Available from https://www.cal-r.org/index.php?id=R-Wsim

Montague, P. R., Dayan, P., and Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience*, 16(5):1936–47.

Moore, J. W., Choi, J. S., and Brunzell, D. H. (1998). Predictive timing under temporal uncertainty: the time derivative model of the conditioned response. In Rosenbaum, D. A. and Collyer, C. E., editors, *Timing of behavior: Neural, Psychological, and Computational Perspectives*, pages 3–34. MIT Press, Cambridge, MA, USA.

Murphy, R. A. and Baker, A. G. (2004). A role for CS-US contingency in Pavlovian conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, 30(3):229.

Nader, K. (2003). Memory traces unbound. *Trends in Neurosciences*, 26(2):65–72.

Nasser, H. M. and Delamater, A. R. (2016). The Determining Conditions for Pavlovian Learning. In Murphy, R. A. and Honey, R. C., editors, *The Wiley Handbook on the Cognitive Neuroscience of Learning*, chapter 2, pages 5–46. John Wiley & Sons, Ltd, Chichester, UK.

Nieoullon, A. (2002). Dopamine and the regulation of cognition and attention. *Progress in Neurobiology*, 67(1):53–83.

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

Niv, Y., Edlund, J. A., Dayan, P., and O'Doherty, J. P. (2012). Neural Prediction Errors Reveal a Risk-Sensitive Reinforcement-Learning Process in the Human Brain. *Journal of Neuroscience*, 32(2):551–562.

Panayi, M. C. and Killcross, S. (2014). Orbitofrontal cortex inactivation impairs between- but not within-session Pavlovian extinction: An associative analysis. *Neurobiology of Learning and Memory*, 108:78–87.

Pavlov, I. P. (1927). *Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex*. Oxford University Press, Oxford, UK.

Pearce, J. M. (1987). A model for stimulus generalization in Pavlovian conditioning. *Psychological Review*, 94(1):61–73.

Pearce, J. M. (1994). Similarity and discrimination: a selective review and a connectionist model. *Psychological Review*, 101(4):587.

Pearce, J. M. and Bouton, M. E. (2001). Theories of associative learning in animals. *Annual Review of Psychology*, 52:111–39.

Pearce, J. M., George, D. N., and Aydin, A. (2002). Summation: Further assessment of a configural theory. *The Quarterly Journal of Experimental Psychology: Section B*, 55(1):61–73.

Pearce, J. M. and Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, 87(6):532–552.

Pearce, J. M. and Redhead, E. S. (1993). The influence of an irrelevant stimulus on two discriminations. *Journal of Experimental Psychology: Animal Behavior Processes*, 19(2):180.

Pearce, J. M. and Wilson, P. N. (1991). Failure of excitatory conditioning to extinguish the influence of a conditioned inhibitor. *Journal of Experimental Psychology: Animal Behavior Processes*, 17(4):519–529.

Rao, R. P. N. and Ballard, D. H. (2004). Probabilistic models of attention based on iconic representations and predictive coding. In Itti, L., Rees, G., and Tsotsos, J., editors, *Neurobiology of attention*, chapter 91, pages 553–561. Elsevier Academic Press, Amsterdam, Netherlands.

Razran, G. H. S. (1939). Studies in configural conditioning: I. Historical and preliminary experimentation. *The Journal of General Psychology*, 21:307–330.

Rescorla, R. A. (1968). Probability of shock in the presence and absence of CS in fear conditioning. *Journal of Comparative and Physiological Psychology*, 66(1):1.

Rescorla, R. A. (1969). Pavlovian conditioned inhibition. *Psychological Bulletin*, 72(2):77.

Rescorla, R. A. (1971a). Summation and retardation tests of latent inhibition. *Journal of Comparative and Physiological Psychology*, 75(1):77.

Rescorla, R. A. (1971b). Variation in the effectiveness of reinforcement and nonreinforcement following prior inhibitory conditioning. *Learning and Motivation*, 2(2):113–123.

Rescorla, R. A. (1972). "Configural" conditioning in discrete-trial bar pressing. *Journal of Comparative and Physiological Psychology*, 79(2):307–17.

Rescorla, R. A. (1980). Simultaneous and successive associations in sensory preconditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, 6(3):207.

Rescorla, R. A. (2003). Protection from extinction. *Learning & Behavior*, 31(2):124–132.

Rescorla, R. A. (2004). Superconditioning from a reduced reinforcer. *Quarterly Journal of Experimental Psychology Section B*, 57(2):133–152.

Rescorla, R. A., Grau, J. W., and Durlach, P. J. (1985). Analysis of the unique cue in configural discriminations. *Journal of Experimental Psychology: Animal Behavior Processes*, 11(3):356–66.

Rescorla, R. A. and Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In Black, A. H. and Prokasy, W. F., editors, *Classical Conditioning II: Current Research and Theory*, chapter 3, pages 64–99. Appleton Century Crofts, New York, USA.

Richards, B. A. (2017). Deep Learning in the Brain. In *Deep Learning and Reinforcement Learning Summer School*, Montreal, Canada.

Saavedra, M. A. (1975). Pavlovian compound conditioning in the rabbit. *Learning and Motivation*, 6(3):314–326.

San-Galli, A., Marchand, A. R., Decorte, L., and Di Scala, G. (2011). Retrospective revaluation and its neural circuit in rats. *Behavioural Brain Research*, 223(2):262–270.

Schrauwen, B., Verstraeten, D., and Van Campenhout, J. (2007). An overview of reservoir computing: theory, applications and implementations. In Verstraeten, D. and Schrauwen, B., editors, *Proceedings of the 15th European Symposium on Artificial Neural Networks, April 25-27*, pages 471–482, ESANN, Bruges, Belgium.

Schultheis, H., Thorwart, A., and Lachnit, H. (2008). Rapid-REM: a MATLAB simulator of the replaced-elements model. *Behavior Research Methods*, 40(2):435–41.

Schultz, W. (2004). Neural coding of basic reward terms of animal learning theory, game theory, microeconomics and behavioural ecology. *Current Opinion in Neurobiology*, 14(2):139–147.

Schultz, W. (2006). Behavioral theories and the neurophysiology of reward. *Annual Review of Psychology*, 57:87–115.

Schultz, W. (2010). Dopamine signals for reward value and risk: basic and recent data. *Behavioral and Brain Functions*, 6(1):24.

Schultz, W., Dayan, P., and Montague, P. (1997). A neural substrate of prediction and reward. *Science*, 275(5306):1593–9.

Smolensky, P. (1987). Information processing in dynamical systems: Foundations of harmony theory. In Rumelhart, D. E., McClelland, J. L., and PDP Research Group, editors, *Parallel Distributed Processing, Explorations in the Microstructure of Cognition: Foundations*, chapter 6, pages 194–281. MIT Press, Cambridge, MA, USA.

Sutton, R. S. and Barto, A. G. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review*, 88(2):135–170.

Sutton, R. S. and Barto, A. G. (1987). A temporal-difference model of classical conditioning. In *Proceedings of the Ninth Annual Conference of the Cognitive Science Society*, pages 355–378, Seattle. Seattle, WA, USA.

Sutton, R. S. and Barto, A. G. (1998). *Reinforcement Learning: An Introduction*, volume 1. MIT press, Cambridge MA, USA.

Symonds, M. and Hall, G. (1995). Perceptual learning in flavor aversion conditioning: Roles of stimulus comparison and latent inhibition of common stimulus elements. *Learning and Motivation*, 26(2):203–219.

Urushihara, K. and Miller, R. R. (2010). Backward blocking in first-order conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, 36(2):281–95.

Wagner, A. R. (1981). SOP: A model of automatic memory processing in animal behavior. In Spear, N. E. and Miller, R. R., editors, *Information processing in animals, memory mechanisms 85*, chapter 1, pages 5–47. Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, USA.

Wagner, A. R. (2003). Context-sensitive elemental theory. *The Quarterly Journal of Experimental Psychology: Section B*, 56(1):7–29.

Wagner, A. R. (2008). Evolution of an elemental theory of Pavlovian conditioning. *Learning & Behavior*, 36(3):253–265.

Wagner, A. R. and Brandon, S. E. (2001). A componential theory of Pavlovian conditioning. In Mowrer, R. R. and Klein, S. B., editors, *Handbook of contemporary learning*, chapter 2, pages 23–64. Lawrence Erlbaum Associates, Inc, Mahwah, NJ, USA.

Wagner, A. R. and Rescorla, R. A. (1972). Inhibition in Pavlovian conditioning: Application of a theory. In Boakes, R. A. and Halliday, M. S., editors, *Inhibition and Learning*, chapter 12, pages 301–336. Academic Press, London, UK.

Ward-Robinson, J. and Hall, G. (1996). Backward sensory preconditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, 22(4):395.

Ward-Robinson, J. and Hall, G. (1999). The role of mediated conditioning in acquired equivalence. *The Quarterly Journal of Experimental Psychology: Section B*, 52(4):335–350.

Wasserman, A. E. and Berglan, L. R. (1998). Backward blocking and recovery from overshadowing in human causal judgement: The role of within-compound associations. *The Quarterly Journal of Experimental Psychology: Section B*, 51(2):121–138.

Whitlow, J. W. and Wagner, A. R. (1972). Negative patterning in classical conditioning: Summation of response tendencies to isolable and configurai components. *Psychonomic Science*, 27(5):299–301.

Whitlow, J. W. and Wagner, A. R. (1984). Memory and habituation. In Peeke H. V. S., P. L., editor, *Habituation, sensitization and behavior*, pages 103–153. Academic Press, New York, USA.

Wilf, H. S. (1990). Generatingfunctionology. *ISBN: 0-12-751956-4*, page 192. Academic Press, Cambridge MA, USA

Wilson, P. N. and Pearce, J. M. (1992). A configural analysis for feature-negative discrimination learning. *Journal of Experimental Psychology: Animal Behavior Processes*, 18(3):265.

Zeithamova, D., Dominick, A. L., and Preston, A. R. (2012). Hippocampal and Ventral Medial Prefrontal Activation during Retrieval-Mediated Learning Supports Novel Inference. *Neuron*, 75(1):168–179.

Zenke, F., Gerstner, W., and Ganguli, S. (2017). The temporal paradox of Hebbian learning and homeostatic plasticity. *Current Opinion in Neurobiology*, 43:166–176.