bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A Real-Time Double Error Correction Model of Associative Learning

Niklas H. Kokkola, Esther Mondragón and Eduardo Alonso
City, University of London
Centre for Computational and Animal Learning Research

In this paper a formal model of associative learning is presented which model incorporates representational and computational mechanisms that, as a coherent corpus, empower it to make accurate predictions of a wide variety of phenomena that so far have eluded a unified account in learning theory. In particular, the Double Error model introduces: 1) a fully-connected network architecture in which stimuli are represented as temporally distributed elements that associate to each other, which naturally implements neutral stimuli associations and mediated learning; 2) a predictor error term within the traditional error correction rule (the double error), which reduces the rate of learning for expected predictors; 3) a revaluation associability rate that operates on the assumption that the outcome predictiveness is tracked over time so that prolonged uncertainty is learned, reducing the levels of attention to initially surprising outcomes; and critically 4) a biologically plausible variable asymptote, which encapsulates the principle of Hebbian learning, leading to stronger associations for similar levels of element activity. The outputs of a set of simulations of the Double Error model are presented along with empirical results from the literature. Finally, the predictive scope of the model is discussed.

Keywords: Associative learning, classical conditioning, computational modelling, real-time double error correction, variable asymptote

Correspondence concerning this article should be addressed to Esther Mondragón, Centre for Computational and Animal Learning Research, 21 Bardwell Road, St Albans, Hertfordshire AL1 1RQ, United Kingdom. E-mail: e.mondragon@cal-r.org

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Associative learning aims at understanding the precise mechanisms by which humans and animals learn to relate events in their environment. Associative learning has been replicated across numerous species and procedures (Hall, 2002; Pearce & Bouton, 2001; Turkkan, 1989) ; its neural correlates have been extensively studied (Gomez *et al*., 2001; Kobayashi & Poo, 2004; Marschner, Kalisch, Vervliet, Vansteenwegen, & Büchel, 2011; Panayi & Killcross, 2014; Roesch, Esber, Li, Daw, & Schoenbaum, 2012); it has proved to be a core learning mechanism in high-order cognitive processes such as judgment of causality and categorization (Shanks, 1995), and rule learning (Murphy, Mondragón, & Murphy, 2008); it underpins a good number of clinical models (Haselgrove & Hogarth, 2011; Schachtman & Reilly, 2011); and its evolutionary origins are beginning to be elucidated (Ginsburg & Jablonka, 2010). It is thus paramount that we develop comprehensive, accurate models of associative learning.

In classical conditioning, a fundamental pillar of associative learning, the repeated co-occurrence of two stimuli (*e.g.*, an odour or tone), S1 and S2, is assumed to result in an association between their internal representations, which entails that the presence of S1 (the conditioned stimulus, CS, or 'predictor') will come to activate the internal representation of a S2 (the predicted stimulus or outcome from now on). When the outcome is a biologically relevant stimulus (unconditioned stimulus, US) able to elicit an unconditioned response (UR) learning 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. The response is assumed to express the strength of the association between the CS and the outcome (Pavlov, 1927), revealing that the outcome is anticipated or predicted by the CS.

In the following sections we proceed with a critical review of different models of classical conditioning structured around four major features in learning theory, namely, the learning rule, the elemental and configural bases of stimulus representation, attentional factors in CS processing, and, finally, the elusive nature of neutral and absent cue learning. We shall proceed with the description of the Double Error model, which we claim provides a formal, coherent corpus for the understanding of classical conditioning. Next, we present a battery of simulations of relevant phenomena, and conclude with a discussion of the main contributions of model.

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The Learning Rule: Error Correction

The acquisition of a conditioned response, assumed to mirror the strength of the CS→US link, usually follows a negatively accelerating monotonically increasing curve over trials (but see Gallistel, Fairhurst, & Balsam, 2004; Glautier, 2013). The associative link's rate of change is thereby proportional to the discrepancy between the expectation and the 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 & 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, Sahota, & Mondragón, 2014; Alonso & Schmajuk, 2012; Balkenius & Morén, 1998; Pearce & 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. The phenomenon of blocking (Amundson & Miller, 2008; Kamin, 1968, 1969; Kohler & Ayres, 1979) showed that when the acquisition training for a cue A is followed by acquisition training with a compound 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 formation of an equivalent link forming 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. Such cue competition during learning is formalized most prominently by the Rescorla-Wagner (RW) model with its 'global' prediction error term (Rescorla & Wagner, 1972), which incorporates within it a sum of the values 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, learning between a CS and US in the RW model is driven by the total discrepancy between the US presence and the expectation elicited for it by all cues, as seen in Equation 1.

$$\Delta V_i = \alpha_i \beta_j x_i (z_j \lambda_j - \sum_l x_l V_l) \,, \tag{1}$$

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

where ΔV_i is the change in associative strength, α_i and β_i are respectively the CS and US intensities, x_i is the activity of the CS being updated, x_l and z_j the activity of any present CS and the US, respectively, and λ_j the US asymptote determining the maximal supported learning. Finally, the summed term is the total extant learning toward the US, calculated as the dot product of the CS activities times their associative strength V_l toward the US. The RW model thus not only accounts for empirical data, but its formulation implied the existence of learning effects not predicted by earlier models or assumed to exist by pre-existing theory, as can be tested in a wide range of simulations (Alonso, Mondragón, & Fernández, 2012; Mondragón, Alonso, Fernández, & Gray, 2013). Primary among these predictions is that if two CSs are independently conditioned to an asymptotic level, then their reinforcement in compound should lead to a decline in their associative strength. This follows from the error term supporting a maximal level of conditioning for all present stimuli on a trial, which is different from the maximal level supported for cues presented in isolation. Such a phenomenon of overprediction has been confirmed (Rescorla, 1970), along with other predictions such as super-conditioning (Rescorla, 1971a), thus strengthening the validity of the model as well as other models relying on a summed error term. Though the RW model has been successful in this regard (Miller, Barnet, & Grahame, 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 the model of Mackintosh (1975). Alternatively, the comparator hypothesis model (Miller & Matzel, 1988; Miller & 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 the comparator CSs. Hence cue competition arises from the interference of other CS-US associations upon the target-US association.

The realization of real-time learning models extending the trial-based RW model (and other error correction models) by making predictions for learning within a trial allows for the modelling of time dependent aspects of learning and temporal relations between stimuli not accounted for by trial-level models. The SB model (Sutton & 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

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

changes in the CS contribution to this node (temporal difference) influenced 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 (Balkenius & 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 & 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 factor 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 (like that in the RW model), the prediction error, for that time-step. Equation 2 is used for calculating this error.

$$\delta^t = (\lambda^t - (\sum_i x_i^{t-1} V_i^t - \gamma \sum_i x_i^t V_i^t)), \tag{2}$$

where λ^t is the US activation at time t, and the term $\sum_i x_i^{t-1} V_i^t$ 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 $\gamma \sum_i x_i^t V_i^t$ is the equivalent for the current time-step. The current time-step 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. 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, Gray, & Alonso, 2013) has produced accounts of phenomena difficult to analyse *a priori*. For instance, higher-order conditioning, wherein reinforcement is observed in the absence of a US through the reinforcement effects of the temporal difference

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

term itself. It accounts for temporal primacy effects – *e.g.*, Kehoe, Schreurs, & Graham, (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 & Morén, 1998) through the aforesaid higher-order conditioning, which in itself results from the temporal-difference term back-propagating to earlier time-points. As the core of the TD model is its distinct error term, variations of TD with differing stimulus representations have been proposed. A complete serial compound model (Moore, Choi, & Brunzell, 1998) postulated that a CS is represented in time by a series of separate units, each of which becomes active following the previous. This extension produced a response curve in time with closer correspondence to empirical fact. Work in timing as well as hippocampal data inspired the micro-stimulus representation of TD (Ludvig, Sutton, Verbeek, & Kehoe, 2009), which instead assumes that the presence of a CS produces a cascade of units to become active in a bell-shaped 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 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 SSCC TD (Mondragón, Gray, Alonso, Bonardi, & Jennings, 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, Bellemare, & Pearson, 2011; Montague, Dayan, & Sejnowski, 1996; Niv, 2009; Niv, Edlund, Dayan, & O'Doherty, 2012; Schultz, 2004, 2006, 2010; Schultz, Dayan, & Montague, 1997).

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 empirical data correlating prediction errors and dopamine signalling, as well as a consideration for parsimony, motivate our use of the summed error correction rule in the Double Error model we are introducing.

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The Nature of the Stimulus Representation: Elemental vs. Configural Models

Causal relations in the world are mostly more complex than the linear CS-US relationship encountered in a standard acquisition protocol, being more appropriately described by non-linear conditional probabilities. Learning models therefore must in some manner involve processes of approximating these conditional 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 discriminative performance of a model have been negative patterning and biconditional discriminations. In negative patterning (NP) individual presentations of two cues A and B are followed by an outcome (a reinforcer), A+/B+, whereas compound presentations of the same cues are not, AB-. Its difficulty and hence importance 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 individually predict the outcome, are presented. Biconditional discriminations involve yet more complex non-linearity. Four cues are presented in pairs, with each individual cue being presented in both a reinforced and nonreinforced compound (AB+/CD+/AC-/BD-). Therefore, simple summation of the individual cues values offers no information for solving the discrimination.

The Rescorla-Wagner (RW) model, as an elemental model, kept with the assumption of Spence, Konorski, and Estes that sets of elements constituting the attributes of individual stimuli enter into associations with the US directly (Harris, 2006; Wagner, 2008). As a result, in its original formulation, RW was unable to account for both negative patterning (Rescorla, Grau, & Durlach, 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 NP, 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 allowed for the discrimination to be solved as the elements common to both A and B acquire superconditioning, super-asymptotic learning (Rescorla, 1971a), 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

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

discriminations within an elemental framework. In the case of more complicated discriminations such as bi- and 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 avoid this issue (Wagner & Rescorla, 1972). That is, it is assumed that the compound is represented by the animal as more than the sum of its parts. Hence NP is produced in a contrary fashion as compared to when common elements are involved, as the configural elements produce inhibitory links toward the outcome, while the unique elements form excitatory associations. While this configural element leads 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, George, & Aydin, 2002) contradicted the RW model. Such evidence 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, while the replaced elements are assumed to represent features of the cue that are uniquely present when the cue is presented alone. The model 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, Thorwart, & Lachnit, 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. Thus, by assuming that different modalities of stimuli undergo different replacement rates depending on their similarity, consistent with psychophysical theory (Glautier, Redhead, Thorwart, & Lachnit, 2010), the model can explain why summation seems to depend on the types of stimuli used (Wagner, 2003).

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

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

observed effects. Both models assume that a redundant cue, X, will facilitate the learning of a discrimination, yet the opposite is often observed (Pearce & 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 impair previous A+/AB- training, contradict this assumption of complete associative reversibility (Wilson & 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 & Wilson, 1991), which seems to require a low replacement rate of elements.

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. This node in return forms associations with the outcome. Its configural stimulus representation, wherein the presentation of a compound activates a configural node representing it, produces many non-linear discriminations simply through learning between a configural node and the outcome not directly interfering with other configurations. Non-linear discriminations are often difficult for animals to learn. 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. In the case of biconditional discriminations, 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 the fact that summation between stimuli is not always being observed. As a result, 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 confirmed that overshadowing produces a larger

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

deficit in responding than external inhibition (Brandon, Vogel, & Wagner, 2000). Finally, evidence has accumulated that familiarity with the constituent stimuli of a discrimination can facilitate the learning of the discrimination (Hall, 1991; Mondragón & Hall, 2002). This phenomenon of perceptual learning (PL) seems to suggest that the animal's perception of the similarity of two stimuli is learned rather than given a priori as through 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 preexposed together, such as elements shared by the preexposed stimuli losing associability.

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.

Harris (2006) and Harris and Livesey (2010) introduced a purely elemental model utilizing a unique attentional process to avoid 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 proposed. 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. An important implication 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

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 discriminative learning by preexposure, *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 & Pearce, 1992), as well as the discrepancy between external inhibition and overshadowing (Brandon *et al.*, 2000). 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 that did. Consequently, 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 complex 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 an elemental approach for the current model.

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Attentional Factors: 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 preexposure and habituation phenomena 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, Schultheis, König, Üngör, & Melchers, 2008). This sets it in opposition to other paradigms of selective attention. 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. 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, 2011). 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 example, it can explain unblocking, wherein the surprising partial omission of reinforcement during a blocking treatment attenuates the blocking of a cue (Dickinson, Hall, & Mackintosh, 1976). It thus avoids the prediction of overprediction pushing the blocked 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 & 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 & Pearce, 1979), wherein reinforcing a CS with a weak US, thus making it a better predictor, hinders subsequent conditioning between the same CS and a stronger US. In this case, it erroneously predicts greater excitatory learning between the previous best predictor of the outcome and

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the outcome. The model also is unable to account for superconditioning (Rescorla, 1971a, 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. Superconditioning is however predicted by Le Pelley's extension of the Mackintosh model (Le Pelley, 2004), owing to the use of a combined error term.

The Hall-Pearce negative transfer effect is easily accounted for by the Pearce and Hall (PH) model (Pearce & Hall, 1980), which postulated that attention rises when the outcome is uncertain. That is, the associability of a CS rises 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, prior training with a weaker US produces a decline in associability which delays 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 predicts a sudden increase in associability when the contingency is changed (*i.e.*, the beginning of acquisition or extinction), with a gradual decline thereafter. Further, a great success of the model has been predicting latent inhibition (LI) (Channell & Hall, 1983), wherein preexposure of a CS attenuates subsequent acquisition training with the same CS. It predicts this effect as arising from the preexposure 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 disentangling increases in associability (speed of learning) from direct reinforcement (extent of learning) produced by prediction errors (Holland & 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 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 Le Pelley's 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

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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.

It is worth noticing that the PH attention rule, when instantiated in a real-time model, can 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. Further, it has been found that even trial-level models based on the PH attention rule can produce Mackintosh-like effects under certain conditions (Le Pelley, Haselgrove, & Esber, 2012).

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 & Schmajuk, 2012). It postulates that the novelty of every stimulus affects its speed of learning towards other stimuli. Further, this novelty is modelled through stimuli learning to predict each other through associative links (expectancy). 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 adjusted through a delta rule. Their associability or associative rate is taken to be very low initially, however if the US expectancy remains high during training (*i.e.*, the animal is unable to learn a given contingency), then the associability of the configural units rises. It hence 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 account for various preexposure 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

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 and Mackintosh (2000) models. In all three models, latent inhibition emerges from the preexposed CS losing novelty by being predicted by the context and itself (unitization). This loss of novelty slows down subsequent acquisition training in a way different from that proposed by the aforementioned PH model and its Hall-Rodriguez extension (Hall & Rodriguez, 2010). These predict that the repeated preexposure of a cue reduces the attention paid to it, thus decreasing its associability. Hall and Rodriguez (2010) further predict the formation of a CS→noUS link, which thereafter interferes with the response elicited by the CS→US link formed in the subsequent acquisition training. This latter mechanism bears a similarity to the explanation offered by the comparator model (Miller & Matzel, 1988), which also stresses processing during retrieval, although in this case, unlike in Hall and Rodriguez's, the effect is specific to a given US. 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 noUS representation or the existence of an extant associative link to it. Hence they avoid predicting inhibitory properties of the preexposed CS towards the US, which have not been found in summation tests (Rescorla, 1971b). Additionally, assuming a noUS representation by implication would lead to generalization of inhibition to another outcome. The evidence seems to contradict such an assumption, as learning with a reinforcer has been demonstrated to be outcome specific, for instance in outcome devaluation experiments (Colwill & Motzkin, 1994).

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 irrelevance. They however are unable to offer a convincing explanation of preexposure effects. Similarly, alternative explanations for unblocking and learned irrelevance can often be produced by error correction models. In the former case by assuming that differences in US presentations imply differences in the asymptote of learning; in the latter case by postulating that uncorrelated CS-US exposure results in inhibitory learning. 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 preexposure. The latter models can also predict the context specificity of LI observed when preexposure is conducted in a context different from that of conditioning.

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Moreover, as above-mentioned, 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 yielding a larger 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 revaluation of persistent uncertainty, to account for learned attentional bias in the Double Error model.

Neutral and Absent-Cue Learning

If an animal is learning the contingencies governing events, does it revaluate 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 data exists, which demonstrates that a present cue can retrieve the representation of another cue and thereby invoke revaluation of the latter's previously formed associations towards a reinforcer (Dickinson, 1996; Holland, 1983; Holland & Forbes, 1982). In addition, neural data indicates that the left hippocampus is involved 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 learning 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 are backward blocking (BB), un-overshadowing/retrospective revaluation (RR) (Le Pelley & McLaren, 2001; Miller & Witnauer, 2016; Urushihara & Miller, 2010), and sensory preconditioning (SPC) (Brogden, 1939; Ward-Robinson & Hall, 1996). Explaining these effects with a simple associative learning rule has been elusive, as they seem to result from opposite learning directions between the cues, retrieved and present.

For instance, in BB and RR designs, a compound of events, AB, is first associated to a motivational outcome. Once this association is established, presentations of one of its components alone, A, follows. While in a backward blocking procedure A is also reinforced, in a RR procedure, subsequent A training is not. Testing of the other element of the compound, B, afterwards, the 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

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

compound and the outcome, results in a reduction in strength of the concomitant stimulus and the outcome, whereas lowering it rises the strength of its associated stimulus.

In a sensory preconditioning procedure, two non-reinforced stimuli, AB, are initially paired. 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 mechanism for this phenomenon relies on the idea of mediated conditioning: during conditioning to A, an associative activation of stimulus B, which was linked to A during the initial compound training, is produced. This associatively activated representation of B acquires associative strength towards the present reinforcer. Of note is that learning between the absent cue, B, and the present outcome in SPC and BB proceeds in opposite directions. That is, while in SPC the absent cue acquires strength, in BB it loses strength.

Different models have been proposed to explain specific sets of mediated phenomena that can be ascribed to a single learning rule, either an increase or a decrease in strength in the same design conditions, but so far, no model can account for all of them. It would thus seem evident that a comprehensive model of learning should be able to predict these apparent contradictory results.

In the McLaren-Mackintosh model (McLaren & Mackintosh, 2000) stimuli are represented by sets of mutually overlapping elements in a connectionist network. Each node in the network 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, as well as acting as a modulator of the associability from that node to other nodes. The internal input is precisely what allows for the retrieval of absent cues and therefore mediated conditioning in the model. McLaren and Mackintosh postulate that in sensory preconditioning, the preexposure of the compound AB results in the formation of bidirectional 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. It is uncertain however whether the model can account for the revaluation effects of backward blocking and retrospective revaluation.

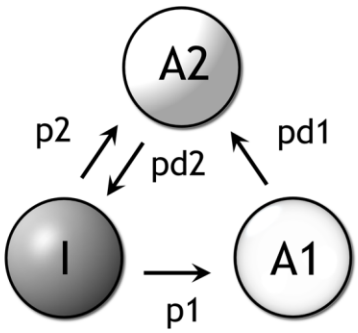
bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 rather than losing it, as its prediction for the outcome would be lower than if it were directly present. That is, the delta error term would still support further excitatory learning, predicting and increase rather than the observed decrease in strength.

Another partially successful model in accounting for mediated phenomena is SOP (Wagner, 1981). SOP postulates that inactive (I) stimulus representations are activated from long-term storage into short-term memory through either stimulus presentation or associative retrieval. The short-term memory system consists of an elemental network of processing units containing elements in primary (A1) and secondary, weaker (A2) states of activation for each stimulus centre/unit, as seen in panel a) of Figure 1. Direct stimulus presentation results in the activation into A1 of a given proportion (p1) of its elements. Active elements (or associatively retrieved elements) decay in time into A2 according to a pd1 proportion. 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. 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 excitatory learning, whereas an over-predicted outcome supports only inhibitory learning. These processes are summarized as learning rules in panel b) of Figure 1.

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

a) SOP activation states



b) SOP learning rules

			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

c) Dickinson & Burke (1996) learning rules

			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

d) Holland (1983) learning rules

			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

Figure 1. a) SOP activation states: I, A1, and A2 are respectively inactive, active, and associatively activated or decayed activation states; 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. b) SOP learning rules. The arrow (→) describes the direction of an excitatory link, whereas (⊣) represents an inhibitory link vector. c) Dickinson & Burke SOP learning rules. d) Holland SOP learning rules.

SOP and its extensions (CSOP (Wagner & Brandon, 2001), and AESOP (Brandon, Vogel, & Wagner, 2003)) assume excitatory conditioning between fully active elements, and the formation of bidirectional inhibitory links between a present CS and an absent outcome (Figure 1, panel b), but do not offer further associative rules. Such rules are introduced by Holland (Holland, 1983), and Dickinson and Burke (Dickinson & Burke, 1996). The former (Figure 1, panel c) accounts for mediated conditioning and predicts that two stimuli in the A2 state undergo inhibitory learning. The latter (Figure 1, panel d) 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. Indeed, experimental data has been collected to support each model (Dwyer, 1999), leaving

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the problem unresolved from the point of view of these two SOP extensions. That is, whether A2-A2 learning should be either excitatory or inhibitory is unclear. Hence, there exist conditions under which A2-A1 learning appears to lead to both excitation and inhibition.

Nonetheless, it is conceivable to explain mediated learning through mechanisms different from alterations in the associative learning rules. For instance, the comparator model produces retrospective revaluation through the competition of associative links during retrieval. That is, the B→US link is relatively larger than the A→US link after the latter has lost associative strength in the second phase. However, a formal analysis of various comparator models has cautioned that such revaluation effects might be harder to reproduce than originally conjectured (Ghirlanda & Ibadullayev, 2015, p. 18). Though the TD model does not presume learning between neutral stimuli, the Predictive Representations (Ludvig & Koop, 2008) extension of it accounts for mediated conditioning by proposing that present cues retrieve representations of stimuli with which they have been previously presented. 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 (2017). It in contrast postulates that the animal re-processes previous learning during the inter-trial interval (ITI), thereby consolidating contradictory learning contingencies. 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 assuming 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.

In summary, if mediated phenomena are to be explained purely through learning rules, what seems to be needed is therefore a learning rule that subsumes both sets of rules. This

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

supra-rule should reverse the direction of learning between present and absent cues in a principled manner depending on the prior reinforcement history. For instance, through a learning process, which can be justified as a form of approximate Bayesian inference. This is precisely the approach we follow in the Double Error model.

To conclude, a plethora of models have been developed to account for wide varieties of classical conditioning phenomena. However, the sets of phenomena explained by models are quite distinct. Latent inhibition is predicted by diverse models, nevertheless the accounts given by the SLGK, McLaren-Mackintosh, and SOP models excel due to their ability to incorporate context modulation. Similarly, solving various non-linear discriminations, such as negative patterning, and modelling exposure effects on stimulus generalization, seem to necessitate a more elaborated stimulus representation than standard elemental and configural approaches. 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 Double Error (DE) model introduced next aims at accounting for all the highlighted classes of learning effects through a general framework of real-time error correction learning, instantiated as an elemental connectionist network. It integrates most prominently a unique second error term denoting the expectancy of an outcome predictor, a revaluation alpha which turns outcome uncertainty into a source of information, and critically, a dynamic asymptote of learning regulated by the similarity/discrepancy in the levels of activation of the elements of the association.

The Double Error Model

The Double Error (DE) model introduced in this paper conceptualizes a formal, computational model of classical conditioning. It is instantiated as a connectionist network, within which nodes representing elements of stimuli enter into associative learning with one another (including between non-US elements, and elements of the same stimulus). It utilises an error correction learning framework, thereby inheriting the properties of extant error correction models. Elements can be unique to a stimulus or shared between two stimuli. 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. Crucially, the asymptote

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

of learning between any two elements measures the similarity of their activity. Hence, elements correlated in their activity patterns form stronger associative links. 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 being revaluated as a source of information.

The time-dependent activity of elements, clustered into 'time-waves', allows the model to predict differential responding early and late during a CS presentation. The learning between motivationally neutral elements explains the influence of learning between non-US stimuli on learning between a stimulus and the US. Thus, the model accounts for preexposure effects such as latent inhibition. The predictor error term, which produces the so called 'double error' learning rule, along with the unique asymptote of learning used, endows the model with the capability of accounting for apparently contradictory mediated learning phenomena (*e.g.*, BB and mediated conditioning). It does so by positing that the crucial factors influencing the direction of mediated learning are the strength of retrieval of the retrieved cue and the prior strength of the link between the retrieved cue and the outcome. Lastly, the attentional processes of the model contribute towards the model's capability of predicting a wide range of phenomena including preexposure, non-linear discriminations, and learned irrelevance effects.

Stimulus Representation

In the DE 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) are probabilistically sampled whenever one of their 'parent' stimuli is active.

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

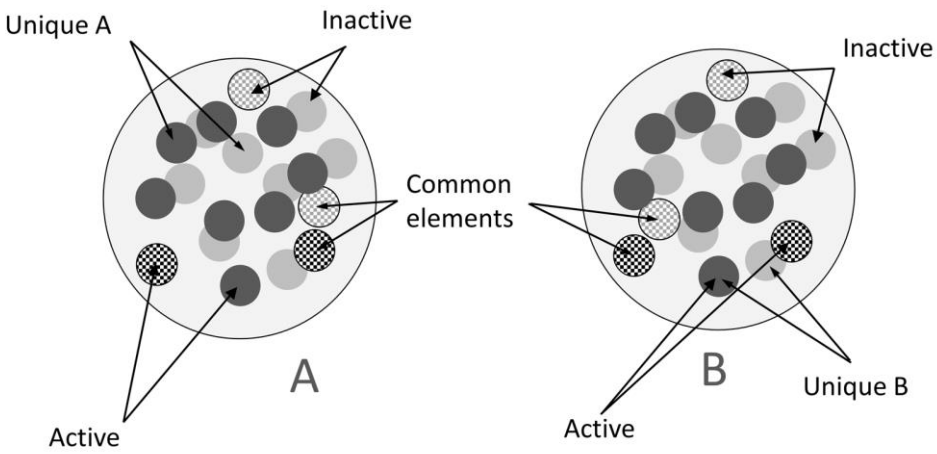


Figure 2. A stimulus consists of elements unique to it, common elements shared with other stimuli, and the elements' respective activation levels.

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 3) of the total stimulus representation, with bidirectional (not necessarily symmetric) links between each pair of elements.

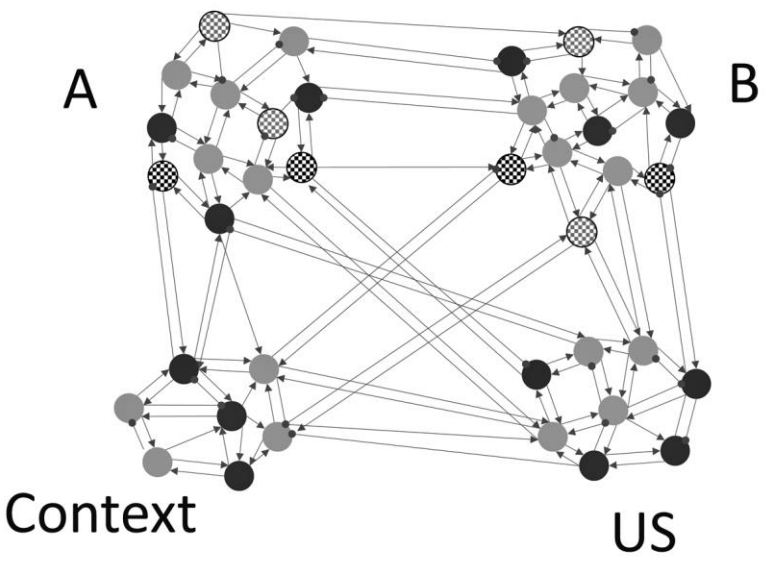


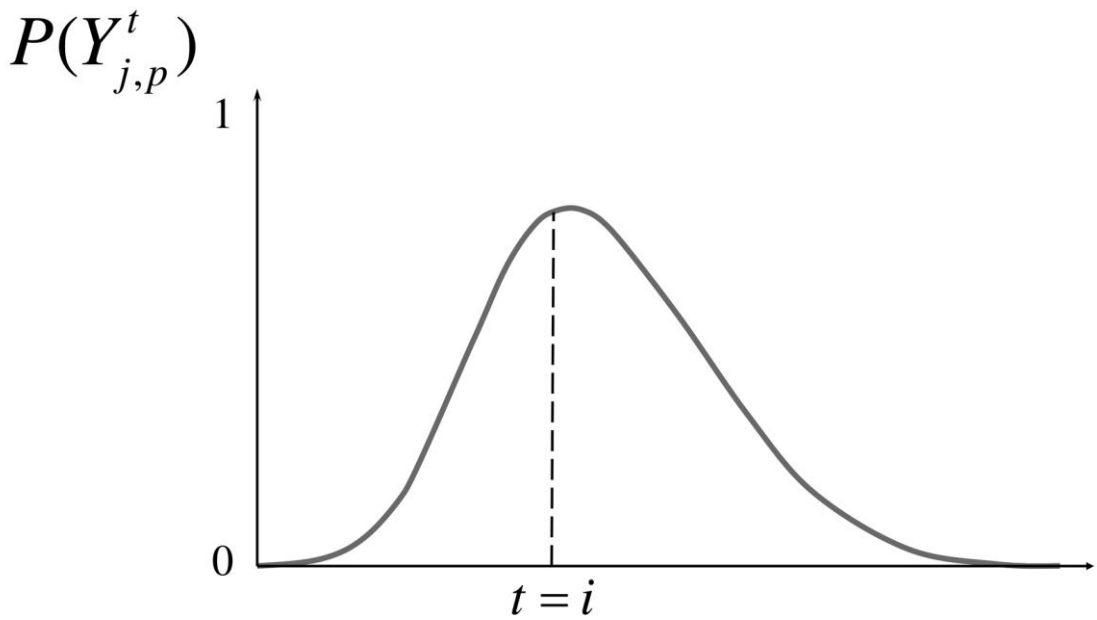
Figure 3. The elemental network structure of the model.

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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, calculated through Equation 3, with $Y_{j,p}^{t}$ denoting the probability that a given element belonging to the time-wave will be directly activated at that moment in time.

$$Y_{j,p}^{t} = \exp(-\frac{(t-i)^2 s}{2c^2}), \tag{3}$$

where $c = \sqrt{\mu/\sigma^2}$, s is a skew parameter which multiplies the enclosed term when $t < i$, and i is the mean of the corresponding time-wave. The curve is not normalized by the usual $\sqrt{2\pi \cdot c}$ 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. Its shape is approximately of the following form (Figure 4):



bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 4. Time-waves are skewed to the right, have a specified peak $t = i$ 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 follow the shape seen in Figure 5.

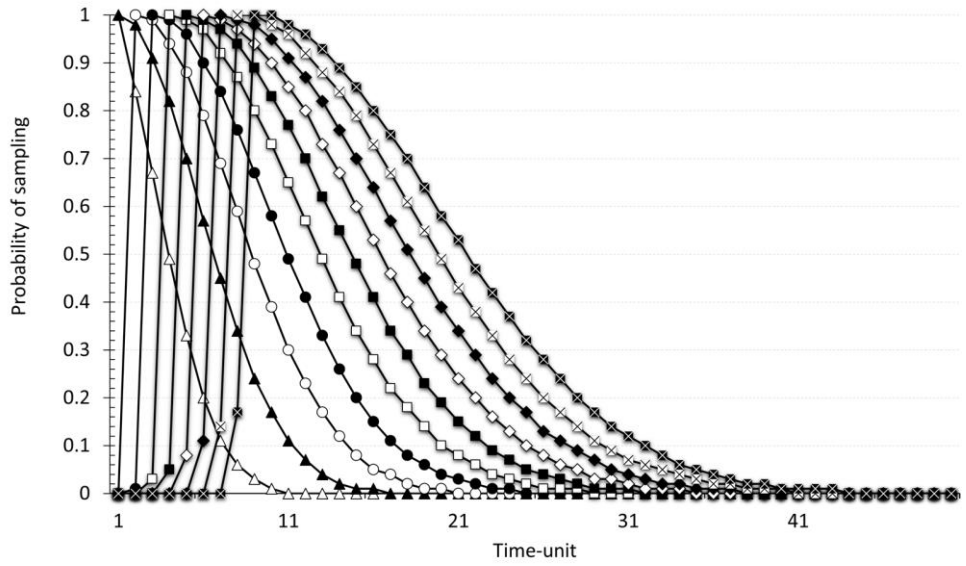


Figure 5. The presence of a stimulus produces a series of time-waves in proportion to its duration. Each time-wave peaks at a subsequent time-point and its standard deviation is proportional to its mean. A given time-wave gives the activation probability of elements 'belonging' to the wave.

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. Thus, earlier time-waves elements can learn differential associations from later ones. Similarly, significant generalization through time occurs as the long-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). The persistence of the time-wave activations also allows for effective trace conditioning. Thus, the DE model does not rely on separate eligibility traces to be able to produce trace conditioning as in other models of learning.

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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. Thus, 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 predicting element j of each stimulus p active at that time-point, that is the link from the predicting element times its overall activation at that moment in time. This is calculated per Equation 4.

$$\hat{Y}^t_{m,o} = \sum_p \sum_j w^t_{j,p \to m,o} \mathcal{A}^t_{j,p} \tag{4}$$

Overall Stimulus Activity

The overall activation of an element m of a stimulus o is taken to be whatever is larger: its direct activation $Y^t_{m,o}$, or its associative activation $\hat{Y}^t_{m,o}$, as given by Equation 5.

$$\mathcal{A}^t_{m,o} = \max(Y^t_{m,o}, \hat{Y}^t_{m,o}) \tag{5}$$

That is, the associative and direct activations of a given element do not compete directly, but rather act complimentarily in terms of invoking activity of an element. Hence predictions formed by one stimulus to another do not inhibit the activation of the predicted stimulus as in

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

models such as SOP (Wagner, 1981), but rather simply reduce the novelty of the predicted stimulus. The motivational value of a US is represented by a US intensity parameter that multiplies both the overall activity of the US elements and the salience of the US. That is, less intense USs have a lower maximal activity level, which further leads to such USs possessing lower learning asymptotes for present CSs.

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 dynamic asymptote, which is a measure of closeness between the activation of the two elements, as well as the associative strength of other cues in the error term. That is, if two elements have similar activities they will undergo more rapid conditioning towards one another and eventually reach higher asymptotic link strength. Hence, two temporally overlapping present stimuli will undergo strong excitatory learning, while a retrieved and a present cue will support a significantly lower maximal level of conditioning. 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'. The power of the dynamic asymptote is that it predicts the same A1→A1 and A2→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→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→A1 conditioning). To reiterate however, the model's behaviour in terms of excitatory and inhibitory learning derives from the error term of the outcome; thus no 'learning rules' are used. 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 have a lower value as compared to A1→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 backward 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

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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. Neural data for stronger retrieved representations tending to be more associable has been gathered by Zeithamova, Dominick, and Preston (Zeithamova, Dominick, & Preston, 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 an inverse 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 calculated through Equation 6 and Equation 7.

$$\widehat{\mathcal{A}}_{j,p} = \begin{cases} 1, & \text{if } Y^t_{j,p} > 0.1 \\ \mathcal{A}_{j,p}, & \text{otherwise} \end{cases} \tag{6}$$

$$\lambda_{j,p \to m,o} = \| \widehat{\mathcal{A}}^t_{j,p}, \widehat{\mathcal{A}}^t_{m,o} \| = \frac{\widehat{\mathcal{A}}^t_{m,o} - | \widehat{\mathcal{A}}^t_{m,o} - \widehat{\mathcal{A}}^t_{j,p} |}{max(\widehat{\mathcal{A}}^t_{m,o}, \widehat{\mathcal{A}}^t_{j,p})} \tag{7}$$

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. This rule is not a strict inverse distance function however, as the absolute value being subtracted from the outcome activity breaks symmetry and allows for differential learning between a present predictor and absent outcome (extinction), and an absent predictor and present outcome (mediated conditioning). As seen in Figure 6, 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-

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

symmetrical, *i.e.*, the outcome activity is more determinant of whether the asymptote is positive or negative.

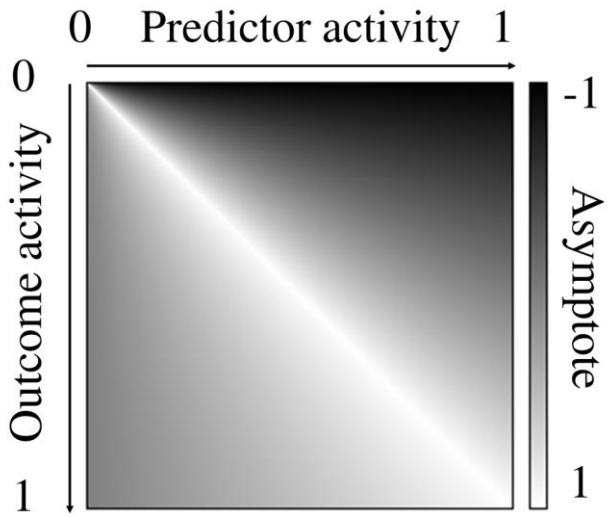


Figure 6. Asymptote of learning as a distance measure of the activities of the predictor and outcome.

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 calculated, per Equation 8, by the discrepancy between the asymptote of learning $\lambda_{j,p\rightarrow m,o}$ and the total prediction for the outcome $\hat{Y}^{t}_{m,o}$.

$$\delta^{t}_{j,p\rightarrow m,o} = \lambda_{j,p\rightarrow m,o} - \hat{Y}^{t}_{m,o} \tag{8}$$

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

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

approximate the activation of other elements at a given moment in time instead of predicting the change in these activities.

Uniquely in the model, per Equation 9, the predicting element itself has an error term denoting how expected it is.

$$\delta^t_{j,p} = |\widehat{\mathcal{A}}^t_{j,p} - \hat{Y}^t_{j,p}| \tag{9}$$

These predictor and outcome error terms are then used in the weight update, Equation 10, for a given element, along with the salience s of the predicting and predicted element, their respective activations $\mathcal{A}$, as well as the adaptive revaluation rate, α_n or α_r (see next section), for a respectively non-reinforced or reinforced outcome:

$$\Delta w^t_{j,p\to m,o} = \delta^t_{j,p\to m,o} \cdot \delta^t_{j,p} \cdot s_{m,o} \cdot s_{j,p} \cdot \mathcal{A}^t_{j,p} \cdot \mathcal{A}^t_{m,o} \cdot \alpha(j,p)^t \tag{10}$$

That is, the direction of learning is determined by the outcome prediction error and its variable asymptote, while the prediction error for the predictor itself, along with other modulating factors (the fixed salience s, adaptive α, and activations $\mathcal{A}$) influence the extent and speed of learning. As such, the novelty of the predicting element plays a crucial role in determining the rate 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 it, is fully accounted for by the model. This crucially allows the DE model to explain preexposure effects such as latent inhibition in a parsimonious manner.

Attentional Modulation: The Stimulus Associability

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

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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. 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. 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 $\mathcal{A}_{j,p}^{t}$ 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, $|\delta_{\text{US/CS}}^{t}|$, as seen in Equation 11.

$$a_{r/n}(j,p)^{t+1} = (1-\widehat{\delta^{t}}_{\text{US/CS}})(1-\rho_{r/n}\mathcal{A}_{j,p}{}^{t})a_{r/n}(j,p)^{t} + \rho_{r/n}\mathcal{A}_{j,p}{}^{t}\,|\delta_{\text{US/CS}}^{t}| \tag{11}$$

Here δ is a moving average of the error terms of respectively all active reinforced or non-reinforced elements, and $\widehat{\delta_{US/CS}^{t}}$ is a decay that 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 learner 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, $|\delta_{\text{US/CS}}^{t}|$, is updated per the formulas of Equation 12 and Equation 13. 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.

$$\delta_{\text{US/CS}}^{t+1} = (1-\rho_{r}\mathcal{A}_{j,p})\widehat{\delta_{\text{US/CS}}^{t-1}} + \rho_{r}\mathcal{A}_{j,p}\,\left|\frac{\sum_{o}\sum_{m}\delta_{m,o}^{t}}{\#\text{allUS/CSelements}}\right| \tag{12}$$

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$$\widehat{\delta_{\text{US/CS}}^{t-1}} = \begin{cases} \delta_{\text{US/CS}}^{t+1}, & \text{if } \delta_{\text{US/CS}}^{t+1} > \xi \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

Hence, the DE model assumes that selective attention, and therefore associability, towards reinforced and non-reinforced outcomes is proportional to the overall uncertainty of these classes of stimuli. The α_r is thus similar to the Pearce-Hall rule but the DE model goes significantly beyond it by postulating a similar process for non-reinforced cues (using α_n), as well as assuming that sustained attention declines if uncertainty remains high. It further predicts that changes in associability induced by uncertainty in one outcome will influence the associability of cues towards other outcomes, both for reinforced and 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 variable associability to a reinforcer decreasing proportionally to learning. However, unlike the Pearce-Hall rate, the effect of the DE model associability is generalized over many trials and its value is dependent upon a moving average of uncertainty that conveys an end of the sustained attention when a sustained consistent high error endures after long training. Thus, the effects of these different approaches to associability only overlap partially.

Computation of Trial Values

To calculate the response elicited by stimuli on a given trial, Equation 14, 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.

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$$R_{\text{Trial}} = \frac{1}{\text{trial length}} \sum_{t=1}^{\text{trial length}} \frac{1}{\#\text{USelements}} \sum_{m=1}^{\#\text{USelements}} \hat{Y}_{m,o}^{t} \tag{14}$$

Model Summary

In summary, 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 whichever is larger: the direct activation or the discounted associative activation.

The model next 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 that 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).

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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.

Results

This section presents a set of simulations aimed at demonstrating that the DE model's unique features enable it to account for a wide variety of the learning effects discussed in the introduction. Simulations were carried out with a universal design simulator (DE Simulator Ver.1, available at https://www.cal-r.org/index.php?id=DE-sim). Results are grouped in three different blocks. The first block tested the contribution of attentional revaluation and the formation of neutral associations to model preexposure effects, namely, latent inhibition effects, the Hall-Pearce effect, learned irrelevance and perceptual learning. The second block of experiments assessed the impact of the model's dynamic asymptote in predicting the individuals' capability to revaluate past associations through retrieved representations of cues. Thus, mediated learning procedures were simulated, specifically, backward blocking, unovershadowing, mediated conditioning in a backward sensory preconditioning procedure, and mediated extinction. Lastly, a third block of simulations examined the competence of the model in coping with complex stimuli and non-linear discriminations such as negative patterning and biconditional discriminations, along with experiments that tested stimulus generalization decrement. This block of simulations was intended to make evident that, with only the assumption of common elements, the DE model can learn to approximate so called configural learning. 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. Thus, when required to parallel experimental results, that is when the analysed response increased inversely to the associative strength, such as in suppression of baseline responding in a conditioned emotional response procedure, or suppression of fluid

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

consumption in a taste aversion learning paradigm, a simulated suppression ratio (r) was computed following (Mondragón *et al.*, 2014), as seen in Equation 15.

$$r = \frac{(\max CR - CR)}{((\max CR - CR) + \max CR)} \tag{15}$$

Here, the baseline normalization constant, max CR rounded to the nearest five, denotes the maximal level of responding elicited during the test condition, that is, normalized to a scale commensurate to the axis units of the empirical data. Values of r closer to 0.5 indicate poor conditioning whereas values near 0 correspond to high levels of conditioning.

The parameters of the simulations are displayed in Table 1 and the design for each experiment is displayed in Table 2. Simulating Experiment 6 requires, due to the large number of cues involved in the design, more RAM than is available on most personal computers. As such, simulating this design needs either the manual disabling of the code for many data arrays normally stored during the running of the simulator or that it be run on a machine with substantially more RAM (*e.g.*, a super-computer).

Preexposure Effects

The following set of experiments involves the presentation of one or multiple stimuli prior to conditioning. Stimulus preexposure treatments are said to entail variations in the processing of the stimulus, which results in a slow rate in subsequent conditioning of the preexposed stimulus to a reinforcer (*e.g.*, latent inhibition effects, Hall-Pearce negative transfer). When the treatment encompasses uncorrelated CS and US presentations conditioning is retarded more than would be expected just by the combined effects of the preexposed stimuli (learned irrelevance), thus suggesting that animals are able to learn an explicit negative correlation between events. Moreover, evidence suggests that this type of learning could also emerge under conditions that may entail a retrospective evaluation of the correlation, that is, under schedules in which the events are experienced separately in time (Baker, Murphy, & Mehta, 2003). Additionally, exposure treatments often result in a reduction of generalization between stimuli (perceptual learning). As such, these phenomena offer the ideal testing ground to demonstrate the model's mechanism of revaluating the

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

associability rate as well as the modulatory effect fostered by neutral cue associations, namely, context→CS and CS→CS associations on subsequent conditioning to a US.

Experiment 1: Latent Inhibition and Context Specificity

An arbiter of a model's ability to fully explain the effect of latent inhibition (LI) is whether the model can also reproduce the fact that LI shows context specificity. Models such as Pearce and Hall (Pearce & Hall, 1980) can account for the deceleration observed in latent inhibition, yet fail to explain the effect that latent inhibition is weaker when subsequent conditioning takes place in a different context than that in which preexposure occurred.

In Experiment 3 (Channell & Hall, 1983) two groups of rats were given preexposure training to a stimulus in a distinctive context (ϕ) 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 in preexposure whereas for the remaining animals conditioning training was given in a different context (ψ)(Group Exposed Different). Two further groups of animals (Group Control Same and Group Control Different) received identical conditioning training but did not received preexposure to the stimulus. Table 2, Experiment 1, shows the design.

The results of this experiment are displayed on the left panel of Figure 7. The conditioning rate was retarded displaying a sigmoidal acquisition shape when the stimulus was preexposed (Group Exposed Same and Group Exposed Different) in comparison to non-preexposed animals. (Group Control Same and Group Control Different). However, this retardation effect was attenuated when conditioning occurred in a context other than that used during preexposure. Thus, Group Exposed Same displayed significantly more latent inhibition, *i.e.*, slower conditioning, than Group Exposed Different, which in return showed slight attenuated learning compared to the control groups.

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

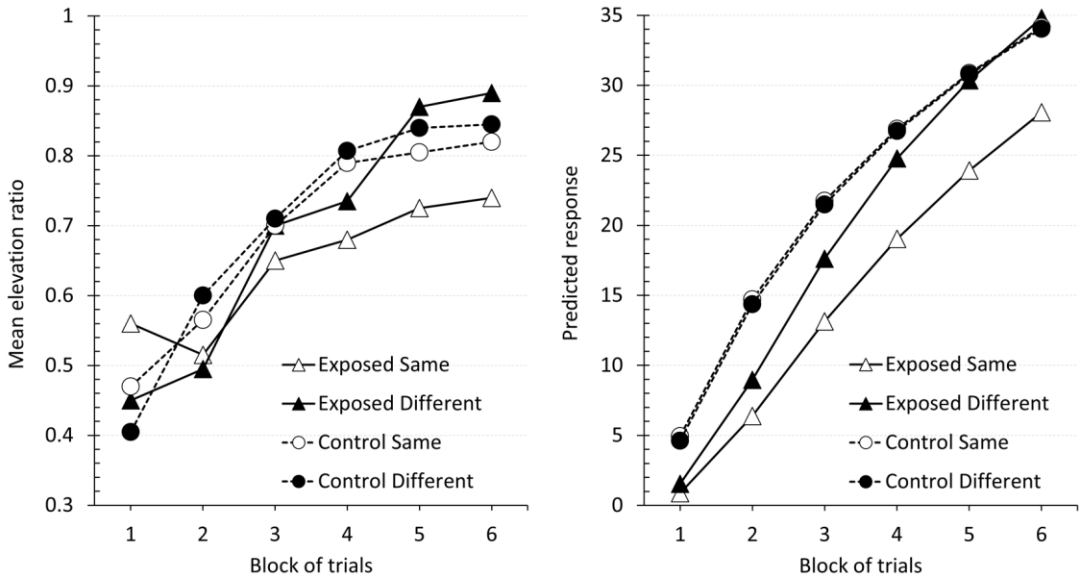


Figure 7: Empirical (original measurement units) and simulated results during the conditioning phase of Experiment 3, Channel and Hall (1983). The left panel is an adaptation of the published data showing acquisition of a CR in groups Exposed Same, Exposed Different, Control Same and Control Different. The right panel displays the corresponding simulated results generated by the DE model.

The results of the corresponding simulation are displayed on the right panel of Figure 7. The duration of the CSs, US and ITI was 5, 1 and 50 time-units, respectively. The values of the remaining parameters were set as per Table 1, Experiment 1. Preexposure consisted of 100 trials in groups Exposed Same and Exposed Different. Group Control Same and Group Control Different received no training in Phase 1. Phase 2 consisted of 66 trials in all four groups, with the Different groups' training programmed to occur in a different context. The results of this simulation (Figure 7, 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 preexposing the stimulus, latent inhibition, although initially evident was considerably reduced with training when the context was changed from conditioning (Group Exposed Different).

During preexposure, 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 DE 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 groups Control Same and Control Different, which have not undergone preexposure and therefore keep their initial associability intact, than in the

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

preexposure groups for which α_r has decayed. In other words, preexposure reduces the associability of cues to future reinforcers as they have a history of not being causative of, or correlated with reinforcement. Hence, the control groups display faster acquisition partly due to the stimulus' 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 preexposure and conditioning occur in the same context, the associations between the context and the stimulus and between the CS elements (unitization), which are formed during preexposure, reduce the speed of the CS→US learning; a decrease which is proportional to the loss of novelty of the CS. As the CS is predicted more strongly in Phase 2 in Group Exposed Same than in Group Exposed Different 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 greater latent inhibition effect when compared to the corresponding non-preexposed control group (Group Control Same). In fact, the degree of latent inhibition is postulated to be proportional to the net effect of unitization, prediction by the context, and loss of selective attention due to lack of reinforcement.

In summary, the DE model's associability accounts for the decelerated learning rate observed in LI. Just as importantly, the predictor error term in the model's learning equation endows it with the ability to predict context modulation of learning, because the CS becomes expected through training in a single context in comparison to when this context is no longer present.

Given that these effects can be disassociated, the model can further predict that latent inhibition would be attenuated (yet not completely abolished) should the preexposure treatment be followed by exposure to the context alone, as this would extinguish the context to CS link. This effect would be slightly mitigated since changes in the preexposure regime would raise the uncertainty of neutral cues, thereby increasing α_n of both the CS and the context. As a consequence of the increase in the neutral cues' associability, stimulus unitization and context to CS learning would presumably be facilitated, thus lessening the attenuation of latent inhibition. Following the same argument, the model can advance a further prediction: preexposing the CS in multiple contexts should result in stronger latent inhibition, as this would reduce the variable associability of the CS, while retaining the strength of the context to CS links.

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Experiment 2: Compound Latent Inhibition

Recent results by Leung, Killcross, and Westbrook (2011) have found evidence of a novel prediction derived from Hall and Rodriguez's model (2010), according to which when a preexposed target stimulus is further exposed in compound with another stimulus latent inhibition accrued by the target is larger than if the target is additionally preexposed in isolation. Hall and Rodriguez's prediction derives from the assumption that preexposure to a CS leads to an association between the CS and an unspecific noUS centre. This association would interfere with the formation of links between that CS and any given US. Unlike isolated CS exposure, additional compound preexposure would result in a reduction of the error due to summation of the two CS→noUS predictions, causing a decline in attention to the target CS. As a consequence, conditioning would proceed more slowly.

Experiment 3 in Leung *et al.* (2011) used an aversive conditioning procedure in rats. In Phase 1, the two groups of animals were presented with non-reinforced random presentations of A, B, and C. Phase 2 consisted of non-reinforced presentations of a compound AB and C. In Phase 3, Group Element received a single reinforced presentation of C (the elemental control), while Group Compound received a reinforced presentation of A (the target). The experiment found that preexposing A in compound with B led to a more pronounced attenuation of subsequent conditioning, that is, to a more robust latent inhibition effect between A and the outcome when compared to stimulus C, which was preexposed in isolation. Figure 8, left panel, shows that animals in Group Element displayed a higher mean percent of freezing, *i.e.*, faster acquisition, than animals in Group Compound during the test phase.

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

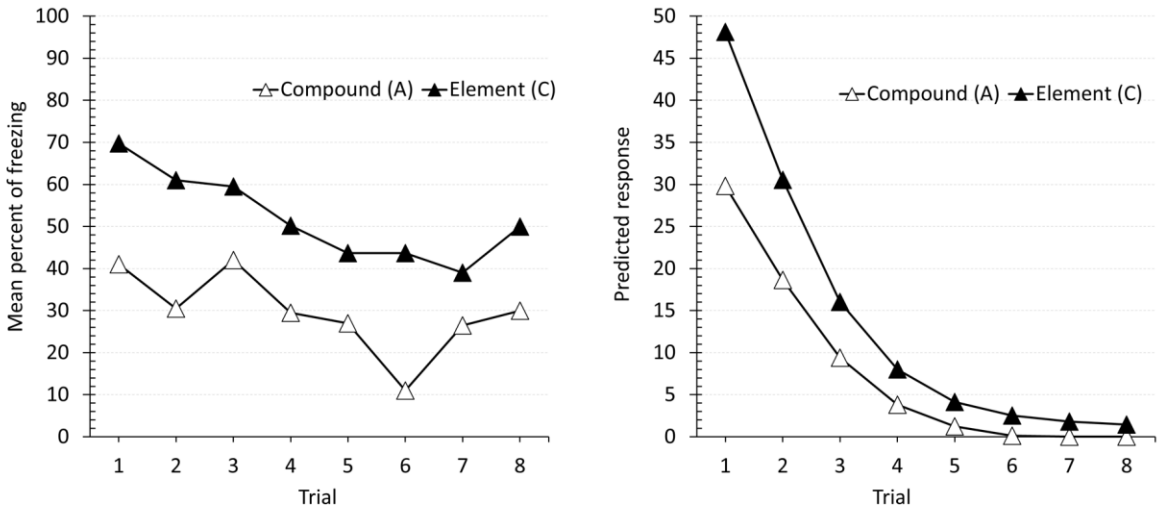


Figure 8: Empirical (original measurement units) and simulated results of Experiment 3 in Leung *et al*. (2011). The left panel is an adaptation of the published data showing the mean percent freezing in the test phase. The right panel displays the corresponding simulated results generated by the DE model.

The temporal parameters used to simulate this experiment were a 5 time-units CS, 1 time-unit US, and 50 time-units ITI. A US intensity of 2.5 was used. The remaining parameters were set to the values in Table 1, Experiment 2. Both groups of the design were programmed to receive 10 randomly presented non-reinforced trials of each stimulus (A, B, C) in Phase 1. In Phase 2, both groups were programmed to receive 10 randomly presented AB and C trials. In Phase 3, Group Compound received a reinforced presentation of stimulus A, while Group Element received a reinforced C trial. The programming of the subsequent test phase presented respectively 8 non-reinforced A and C trials for groups Compound and Element. The full experimental design is presented in Table 2.

The simulation reproduced the pattern observed in the experiment. Figure 8, right panel, shows that the predicted responding strength to A in Group Compound was lower than to C in Group Element during the test trials of the last phase, thereby displaying more latent inhibition.

The DE model's account of this effect is simple and does not require extra assumptions as those in (Hall & Rodriguez, 2010). Neutral associations between the constituent elements of the compound AB are formed during preexposure, so that the two CSs become predictors of one another. This leads to A retrieving B elements during conditioning in Group Compound,

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

which in return produces a second-order prediction for A elements. Thus, A in Group Compound is predicted by both the context and the retrieved B. Since the DE model assumes a predictor error term within the error determining the change in the associative weight of a stimulus, it predicts that A would condition at a slower rate in comparison to the speed of acquisition of C in the Group Element, which is predicted only by the context.

This is of considerable interest, as it implies that the DE model does not require to postulate acquisition of an arbitrary and unspecific CS→noUS association during preexposure (Hall & Rodriguez, 2010) to account for these effects. The DE model's unique learning rule, which incorporates the unexpectedness of a predictor (in this case the CS) in the learning equation, thus modulating the rate of conditioning between a predictor and the outcome, suffices.

The suggested mechanism permits the model to make a further prediction: the potentiation of the LI effect observed in this experiment should be considerably attenuated should B undergo a series of non-reinforced presentations after being paired with stimulus A, but before the reinforcement of A. This treatment would weaken the associative connection between B and A, thereby reducing the proportion of A elements cued by B during conditioning, accelerating the rate of acquisition.

Experiment 3: Hall-Pearce Effect

The acquisition of a CS→US link does not always proceed monotonically in relation to only the pre-existent associative strength. For instance, preexposing a CS with a weak US has been proved to attenuate subsequent acquisition with a stronger US (the Hall-Pearce effect). The error correction and revaluation alpha processes of the DE model can reproduce this observation.

In Hall and Pearce (1979) Experiment 2 two groups of rats received presentations of a tone (Group Tone-shock) or a light (Group Light-shock) followed by a weak shock. A third group, Group Tone-alone, received non-reinforced presentations of the tone in isolation. In Phase 2, all three groups of rats received presentations of the tone followed by a strong shock. The design of this experiment is displayed in Table 2. The experiment found that reinforcing a CS with a weak shock retarded subsequent acquisition towards a more intense shock. This attenuation of learning was however less pronounced than that produced by preexposure of a CS. The results of this experiment are displayed in the left panel of Figure 9. Group Tone-

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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-units CS, 1 time-unit US, and 50 time-units ITI. Aside from these values, the parameters corresponded to those in Table 1. In Phase 1, each trial-type was programmed to occur 66 times with a US intensity of 0.1. Likewise, Phase 2 consisted of 66 reinforced trials, with a US intensity of 1.0.

The simulated results (Figure 9, right panel) paralleled empirical data. 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.

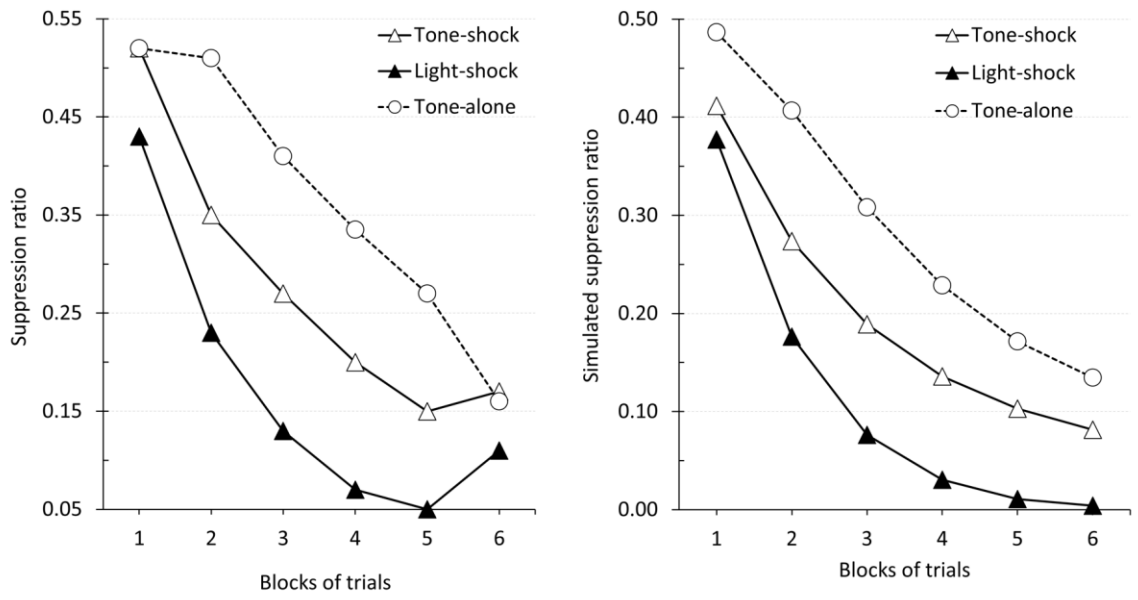


Figure 9: Empirical (original measurement units) and simulated results of Experiment 1 in Hall and Pearce (1979). The left panel is an adaptation of the published data showing the acquisition of a CR in groups Tone-shock, Light-shock, and Tone-alone. The right panel displays the corresponding simulated suppression ratios generated by the DE model.

The DE model accounts for this result in the following manner: in Phase 1, exposure to the target stimulus results in a loss of the tone's novelty in Group Tone-alone and Group Tone-shock. This loss of novelty has a two-folded effect. On the one hand, the tone is predicted by

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the context at the time of conditioning, which reduces the predictor error term and therefore, the rate of conditioning to US; on the other, there is a decline in the tone's associability rate to the reinforcer. Due to the presence of the shock in Group Tone-shock, this decline in the tone's associability rate is less pronounced than in Group Tone-alone. In contrast, in Group Light-shock the tone has not been preexposed by itself and therefore is neither predicted nor affected by a loss of associability.

Since the effect of the delay in conditioning under these conditions is, according to the DE model, a product of a decrease in the error and a loss of the stimulus associability, one could predict 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.

Experiment 4: 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 & Hall, 1996; Mackintosh, 1973). 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

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 10 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 11 left 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.

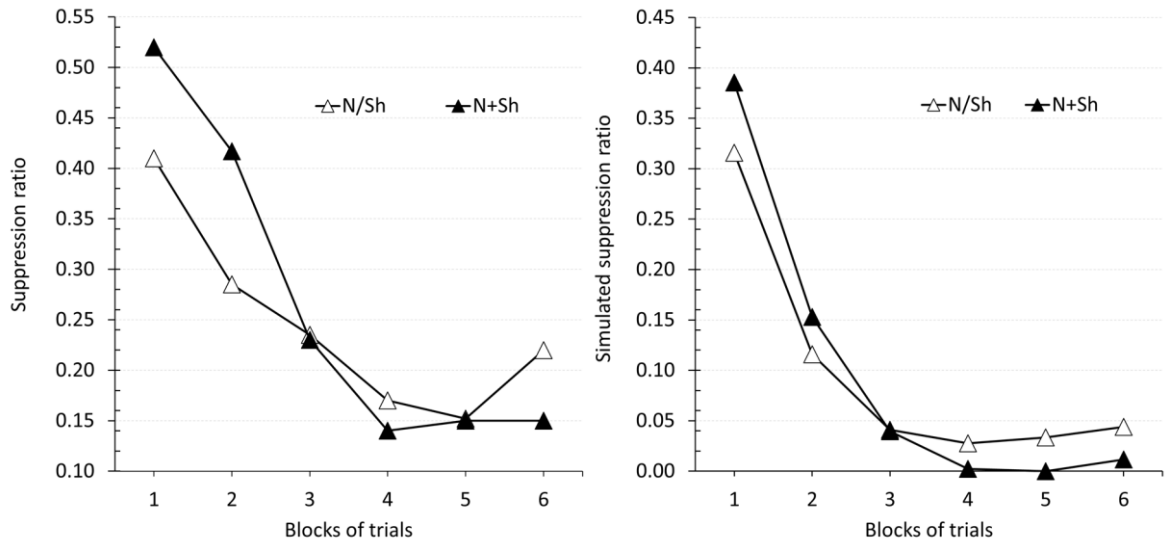


Figure 10: Empirical (original measurement units) and simulated results of Experiment 1a in Baker *et al.* (2003). The left panel is an adaptation of the published data showing suppression ratios for the retardation test trials to

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the noise in Phase 3. The right panel displays the corresponding simulated suppression ratios generated by the DE model.

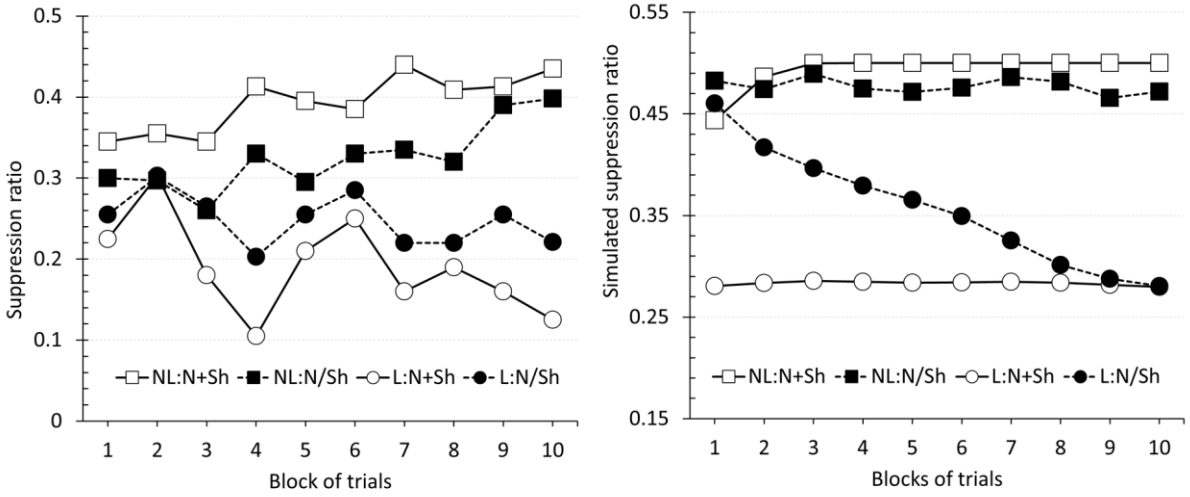


Figure 11: Empirical (original measurement units) and simulated results of Experiment 1b in Baker *et al.* (2003). The left panel is an adaptation of the published data showing suppression ratios for the summation test trials. The right panel displays the corresponding simulated suppression ratios generated by the DE model.

For the simulation, temporal parameters consisted of a 5 time-units CS, 1 time-unit US, and 50 time-units ITI. The remaining parameters were set to the values in Table 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 16 trials. Half of these US presentations occurred before N, and half following it. Group N+Sh received a block of 24 N trials followed by a block of 24 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 2.

The simulation results match empirical data. In the retardation test (Figure 10, right panel), Group N+Sh showed a lower suppression of responding than Group N/Sh.

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Results for the summation test (Figure 11, right 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. 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.

Experiment 5: Perceptual Learning

Perceptual learning (PL) is an effect whereby exposure of stimuli reduces generalization between them or, equivalently, improves subsequent discrimination. The phenomenon is highly relevant in learning theory because it is in apparent conflict with latent inhibition: whereas stimulus exposure delays acquisition, it also facilitates discrimination learning. Amelioration in discrimination has been found to be more profound when the preexposed cues are intermixed (strictly alternated), as compared to being presented in blocks of trials (*e.g.*, Hall & Honey, 1989); Mackintosh, Kaye, & Bennett, 1991); Mondragón & Murphy, 2010); Symonds & Hall, 1995).

Blair & Hall (2003), Experiment 1a, employed a within-subjects design to further control 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 a generalization test. In Phase 1, rats received non-reinforced exposure to three flavours,

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 in the second half. This schedule was counterbalanced across animals, such that half of the animals experienced first a block of CX and then the alternated AX and BX trials. In Phase 2, AX trials were followed by a LiCl injection to induce flavour aversion to AX. In the test phase thereafter (Figure 12, left panel), consumption of BX was higher than consumption of CX, implying that the aversive learning to AX 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 2.

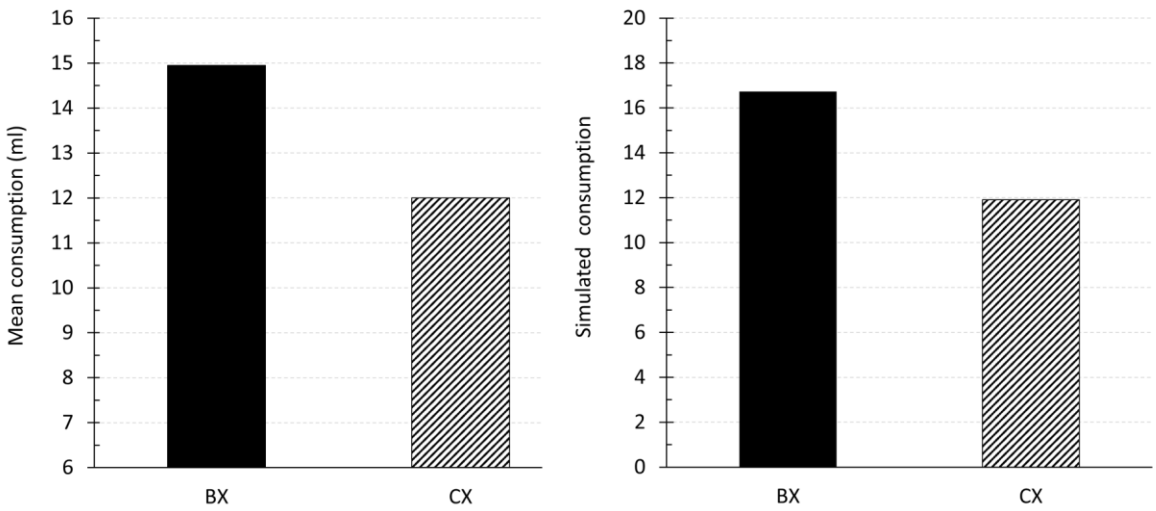


Figure 12: Empirical (original measurement units) and simulated results of Experiment 1a in Blair and Hall (2003). The left panel is an adaptation of the published data showing mean fluid consumption. The right panel displays the corresponding simulated mean consumption for BX and CX test trials.

A simulation of this experiment was conducted with the following temporal parameters: 5 time-units compound CSs, 1 time-unit US, and 75 time-units ITI. All other parameters are shown in Table 1. During Phase 1, each compound was presented 10 times in accordance with the experimental schedule. Conditioning in Phase 2 consisted of 10 trials. In Phase 3, 2 generalization trials to BX randomly interspersed with 2 CX trials were programmed.

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The simulated results (Figure 12, 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 presentations.

According to the DE model, intermixed AX and BX preexposure results in: 1) slightly higher associability in C than B, such that both neutral associations and subsequent reinforcer associations develop at a higher rate; and 2) a stronger A→C association than the A→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→A than C→A links. Thus, during the test phase, the associative chain B→A→US is weaker in comparison to C→A→US, contributing towards the disparity between the intermixed and blocked conditions.

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 may be associatively retrieved yet physically absent. It accomplishes this through both its fully connected network 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 that have given origin to conflicting models.

For instance, in a retrospective revaluation experimental setting, two stimuli, a paired-CS and a target-CS, undergo reinforced training in compound. After this training is completed, the paired-CS is further trained (either reinforced or non-reinforced) 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 unovershadowing (often referred to as retrospective revaluation) are exemplary cases of mediated phenomena. In a backward blocking procedure, following reinforced compound training, the paired cue is subsequently trained with the same

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

reinforcer. As a result, the target cue losses some of its initial strength. In an unovershadowing design, the paired cue is presented in extinction instead. Following the extinction trials, the associative strength of the non-present target is found to increase when subsequently tested. At a face value, these results seem to suggest an inverse (or inhibitory) relationship between active and retrieved cues.

Formally identical treatments such as sensory preconditioning (SPC) and mediated extinction (ME), in which the initial compound training is giving in the absence of a reinforcer, have produced results that opposed the hypothesis above. That is, they suggest a direct or excitatory connection between retrieved and present cues. 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 ME 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 compared to that attained earlier is observed. This result contradicts the predictions of theoretical approaches that are able to account for retrospective revaluation experiments, unovershadowing designs in particular, in which an increase in strength is obtained instead.

Experiment 6: Unovershadowing and Backward Blocking

In Experiment 3 (Le Pelley & McLaren, 2001), mediated learning effects were studied in a causal judgement task using human participants, comprised of a series of mediated learning conditions and controls (see Table 2 for details). All participants received the whole set of conditions. In condition A2-A2, Phase 1 consisted of non-reinforced presentations of compound AB, followed by reinforced presentations of C. In Phase 2 the subjects received non-reinforced presentations of a compound AC. In condition A2-A1, Phase 1 consisted of non-reinforced presentations of a compound DE followed by non-reinforced F presentations. Phase 2 consisted of reinforced DF presentations. In condition Control non-reinforced presentations of a compound GH were followed by reinforced presentations of I in Phase 1, whereas in Phase 2 non-reinforced GJ trials were given. The Unovershadowing condition consisted of reinforced KL presentations in Phase 1 followed by non-reinforced presentations of K in Phase 2. In the Backward Blocking condition reinforced MN presentations were given in Phase 1, and Phase 2 consisted of reinforced presentations of M. In the RR Control condition subjects received reinforced OP trials in Phase 1, and O was partially reinforced in

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Phase 2. In condition Fillers, Phase 1 consisted of non-reinforced Q trials, and reinforced QR trials were given in Phase 2.

The most interesting results of this experiment (see Figure 13, left panel) show that responding to the target cue in the Unovershadowing condition L was higher than that of the Backward Blocking condition N. The difference in ratings between L and P, indicative of unovershadowing, was larger than that between N and P. However, no differences were found in the ratings between P and N, failing to replicate backward blocking, found in their previous experiments. Additionally, no differences were found either between the target cues in the A2-A2, the A2-A1 and the Control conditions (cues B, E, and H respectively) showing no evidence of mediated learning. These results pose a challenge to models such as Holland (1983) and SOP, which cannot predict unovershadowing or backward blocking. These can be considered as post-acquisition effects resulting from memory interference rather than new learning (McLaren, 1993).

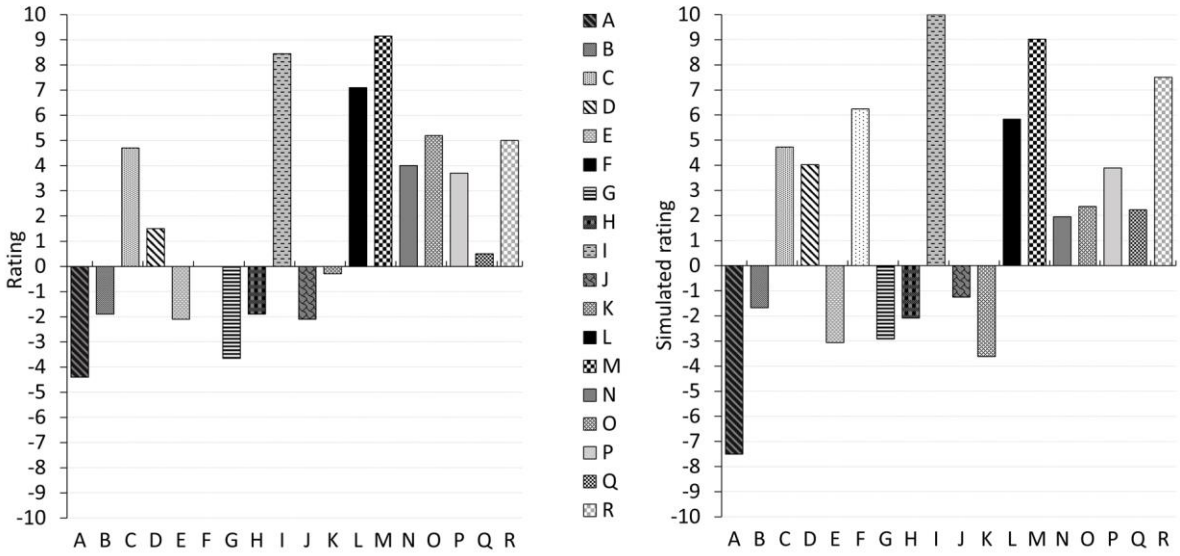


Figure 13: Empirical (original measurement units) and simulated results of Experiment 3 in Le Pelley and McLaren (2001). The left panel is an adaptation of the published data showing mean ratings in a food allergic reaction task during test. Positive ratings indicate high likelihood of a given food resulting in an allergic reaction. Negative ratings indicate that the food would prevent allergic reaction. The right panel displays the corresponding simulated ratings for each cue.

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The temporal parameters of the simulation used a 1 time-unit CS, 1 time-unit US, and 20 time-units ITI. All other parameters were set as per Table 1. Phase 1 of the experiment was programmed to consist of 8 trials of each trial-type. Phase 2 was conducted with 8 trials for each condition. In Phase 3 each cue was tested in a single trial in a random order. To calculate simulated ratings, the weights were normalized such that the highest weight corresponded to a rating of 10.

The simulations matched the empirical data closely (Figure 13, right panel). Crucially, simulated ratings to the Unovershadowing target cue L were higher than those observed for the Backward Blocking condition N. Simulated data showed differences in the ratings between P and N, thus exhibiting backward blocking, and between the target cues in the A2-A2, the A2-A1 and the Control conditions (cues B, E, and H, respectively) which were however rated negatively as in the published experiment.

The DE model predicts that the associative links of cues B, E and H have such low ratings towards the outcome, because during Phase 1 they are presented in the context of reinforced trials of other conditions, resulting in the context acquiring a mild excitatory strength thus fostering inhibition to B, E and H, therefore counteracting mediated acquisition. Furthermore, the low amount of trials in Phase 1 does not allow for significant within-compound links to form between the CS compounds, which impedes Phase 2 mediated learning by leading to a lower level of retrieval of the target cue.

In the Unovershadowing Condition, reinforcement of K and L is followed by non-reinforcement of K. This leads to excitatory revaluation of the L link, as the relative probability of K being a predictor of the outcome diminishes. Specifically, the links K→L and K→+ from Phase 1 of the experiment result in K retrieving L and the US to a similar level of activation in Phase 2. As the K→+ link extinguishes over these presentations, and as the asymptote in the model supports a high level of learning from L to the US due to their similar level of activity, the net result is a growth of the link L→+. In the case of Backward Blocking, a similar revaluation takes place, but in the opposite direction. The link M→N that forms in Phase 1 of the design leads to M retrieving N in Phase 2. As the asymptote of learning in the model between the retrieved N and the present US would be lower than when both cues were present in Phase 1 (due to the disparity in the level of activations), this would result in a decrease of the strength of the association between N and the US.

Further, since this decrease is dependent on how strongly N is retrieved by M, as well as on the extent to which N contributes towards predicting the outcome, the model predicts that

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

a larger number of compound trials in the first phase should result in more backward blocking being observed.

Experiments 7a and 7b: Backward Sensory Preconditioning (BSP) and Mediated Extinction (ME)

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 (Ward-Robinson & Hall, 1996) aimed to uncover the mechanism underlying SPC, that is, whether sensory preconditioning operated by means of an associative chain X→ A→ US or whether a direct link X→US was formed through mediated learning. To do so, they employed backward serial presentations of the compound stimuli during the initial training, that is A→X, with the intent of preventing the chain of 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 although weak, an excitatory link would indeed be formed between X and A during backward conditioning. Thus, during test, X would still be able to associatively activate the chain X→A→US. However, during conditioning of A, the target X would also be active, and mediated conditioning A→US would occur. Despite that the DE Model advances two sources able to produce the effect, mediated conditioning would nonetheless be weak given that the discrepancy between the levels of activation of the stimuli involved would bear a low asymptote of learning. The conjunction of both mechanisms would therefore be needed to produce the effect.

In Experiment 2 (Ward-Robinson & Hall, 1996) two groups of rats received random presentations of two serial compound trial-types, A followed by X, and B followed by Y in Phase 1. In Phase 2, presentations of A were followed by an outcome whereas presentations of B were not. Thus, B and its paired Y acted as a within-subjects control for BSP. In Phase 3, Group Ext received non-reinforced presentations of each A and B, intended to extinguish their association between A and the US, while Group VI received none. The test phase, Phase 4, consisted of X and Y trials for both groups (see Table 2). If the association between A and the US were critical to the effect, BSP would only to be expected in Group VI.

Results (Figure 14, left panel) showed that in Group VI, suppression to X, paired with A, which was reinforced in Phase 2, was greater than to Y, paired with B, which in turn was

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

presented without reinforcement. That is, a BSP effect was shown in Group VI but not in Group Ext, for which extinction training following conditioning was given. In this group, suppression to X did not reliably differ, 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 BSP effect as observed in Group VI.

The temporal parameters used for this simulation were a 5 time-units CS, 1 time-unit US, and 50 time-units ITI, with remaining parameters presented in Table 1. In Phase 1, all groups were programmed to receive 12 random presentations of serial-compounds A→X and B→Y. In Phase 2, both groups received 4 presentations each of trace conditioning A→+ and non-reinforced B trials, presented in the following order MNNM. In Phase 3, 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 (Figure 14, right panel) replicated the empirical pattern of responses. In the test phase, suppression to X in Group VI was greater than suppression to Y. However, unlike the empirical results, a smaller but clear difference between X and Y was also present in Group Ext. This difference could just be attributed to a more complete extinction of the association between A and the US during the simulated Phase 3 in comparison to the observed empirical levels, in conjunction to a lesser stimulus generalization. It is worth noticing that, despite not being significant, a similar tendency can be observed in the empirical data in the last test trial, after further extinction occurred.

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

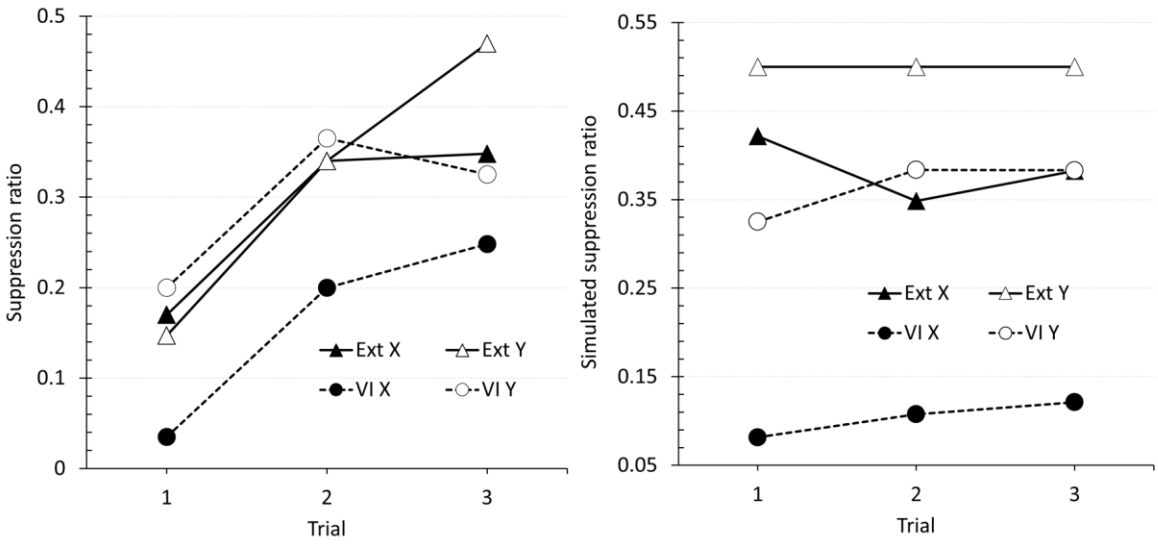


Figure 14: Empirical (original measurement units) and simulated results of Experiment 2 in Ward-Robinson and Hall (1996). The left panel is an adaptation of the published data showing mean suppression ratios to X and Y during the test phase for Group VI and Group Ext. The right panel displays the corresponding simulated suppression ratios.

Ward-Robinson and Hall's Experiment 3 (Ward-Robinson & Hall, 1996) tested (forward) mediated extinction in a within-subjects design that paralleled the one above (see Table 1). Animals received serial A→X and B→Y trials in Phase 1, followed by conditioning to X in Phase 2 in which Y was also presented but in extinction. Phase 3 consisted of presentations of A in extinction and finally, during Phase 4, test trials to X and Y were given.

If mediated extinction occurred during Phase 3, conditioning to X associatively retrieved by A should be reduced in comparison to that of Y. Results of this experiment (Figure 15, left panel) showed that indeed this was the case. During test, suppression to X was lower than to Y.

The parameters and design used in this simulation are displayed in Table 1. The simulation used a US intensity of 2.0. During Phase 1, animals received identical training to those in Experiment 7a. In Phase 2 however, X and Y (rather than A and B) were both reinforced. In Phase 3, A was extinguished and responding to X and Y tested in Phase 4.

The simulated 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

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

strength attained by Y whose paired stimulus B did not undergo extinction (Figure 15, right panel).

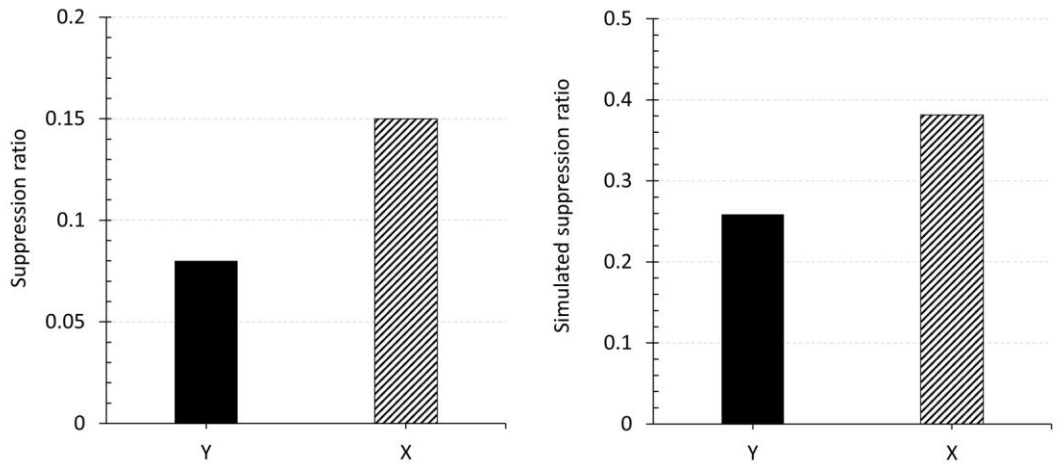


Figure 15: Empirical (original measurement units) and simulated results of Experiment 3 in Ward-Robinson and Hall (1996). The left panel is an adaptation of the published data showing mean suppression ratios to X and Y during the test phase. The right panel displays the corresponding simulated suppression ratios.

The DE model predicts that in Experiment 7a, presentations of A→X and B→Y in Phase 1 would result in bidirectional excitatory learning between the two CSs of the serial compound, as the CS that is presented first is presumed to leave a persistent memory trace even after its offset and excitation occurs whenever two similarly active elements co-occur. In the subsequent conditioning to A in Phase 2, A retrieves X. Although the lesser activation of X implies a lower asymptote of learning towards the present US, this nevertheless increases the associative strength of X. The extinction A trials in Group Ext lead to: (1) a decrease in the association between A and X; (2) a reduction in the association between A and the reinforcer; and (3) a mediated loss in the US expectation elicited by X, as A retrieves the US more strongly than X. This discrepancy in the levels of activation would foster the development of inhibitory learning between the weakly retrieved X and the strongly retrieved US on the basis of the dynamic conceptualization of the asymptote in the DE model. The absence of extinction trials in Group VI would preclude the effects described above from occurring, and thus X would still sustain the level of prediction of the US attained during conditioning to A. However, since the DE model assumes that bidirectional links will be

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

formed despite the serial stimulus presentation, the contribution of the associative chain X→A→US cannot be diminished. Nonetheless, the extinction treatment in Group Ext would prevent A eliciting the US representation, thus countering the BSP effect. Further, if the model is accurate, it would imply that should the first and second phases of the experiment be 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→X link to a level rivalling the more speedily acquired A→US association, while shortening the second phase would achieve a similar result by weakening the A→US association. This would subsequently lead to X and the US being retrieved to an equally active level on the A- trials, in which case the dynamic asymptote of the model would predict a strong excitatory association from X to the US. The mediated extinction hypothesis was further validated by the results of Experiment 7b. The experiment consisted of an equivalent Phase 1 treatment as experiment 7a, and hence the model predicts equivalent X→A and Y→B links forming in this phase. During the second phase, the DE model predicts that, due to the aforesaid links, both X and Y would strongly retrieve a representation of A and B respectively. The high level of activation of these representations would produce a high asymptote of learning from these cues to the present US, and thus would undergo mediated conditioning. During the crucial third phase, non-reinforced presentations of A would retrieve X and the US, but the retrieved representation of the US would be weaker than that of X (reflecting the link strength from A to these cues). Therefore, the asymptote of X toward the US would be smaller than its prediction, and hence the error of X would become negative, resulting in X losing strength (mediated extinction). Additionally, the associative chain X→A→US would weaken due to extinction of the A→US link in Phase 3. Hence, in the test phase, X would elicit less responding than Y, which did not undergo such mediated extinction.

Non-linear Discriminations

Non-linear discriminations are those in which a linear summation of the associative strengths of the constituent CSs is insufficient to accurately predict when reinforcement will or will not occur. Solving these discriminations therefore requires the use of additional representational, learning, and attentional processes to introduce non-linearity in the system. Despite its elemental nature, the DE model is able to solve complex non-linear discriminations primarily through its assumption of elements being able to be activated by

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

multiple stimuli. This effect is accentuated by both the alpha and the 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 partial reinforcement, eventually triggers a decay mechanism of the revaluation alpha that disproportionately affects common elements. The latter operates through the network of neutral element associations to reduce the learning rate of cues that are predicted well by other cues, thereby allowing stimuli that are predictive of reinforcement or non-reinforcement to learn faster.

Experiment 8: Negative Patterning (NP)

Whitlow and Wagner's (1972) Experiment 1, studied a NP discrimination in eyeblink conditioning. In Phase 1, rabbits were presented with reinforced presentations of A, B, and C. Phase 2 consisted of random presentations of reinforced stimuli A and B, as well as the non-reinforced compound AB trials. In Phase 3, training continued as in Phase 2 with the addition of reinforced C trials. Finally, in Phase 4 isolated stimulus presentations and all combinations of the three cues were tested in a random order, that is it consisted of A, B, C, AB, AC, and BC presented in extinction. Stimulus C was used to assess the contribution of simple summation in the responding elicited by a compound in comparison to putative configural components emerging from training a stimulus compound.

Results of this experiment showed that the animals learned to discriminate between the stimuli and the compound AB (Figure 16, left panel), withholding responding on AB trials. Also, test showed that this suppression of the response was only evident for AB trials, that is, the animal's response to the other compounds tested, *i.e.*, AC and BC was sustained (Figure 17, left panel).

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

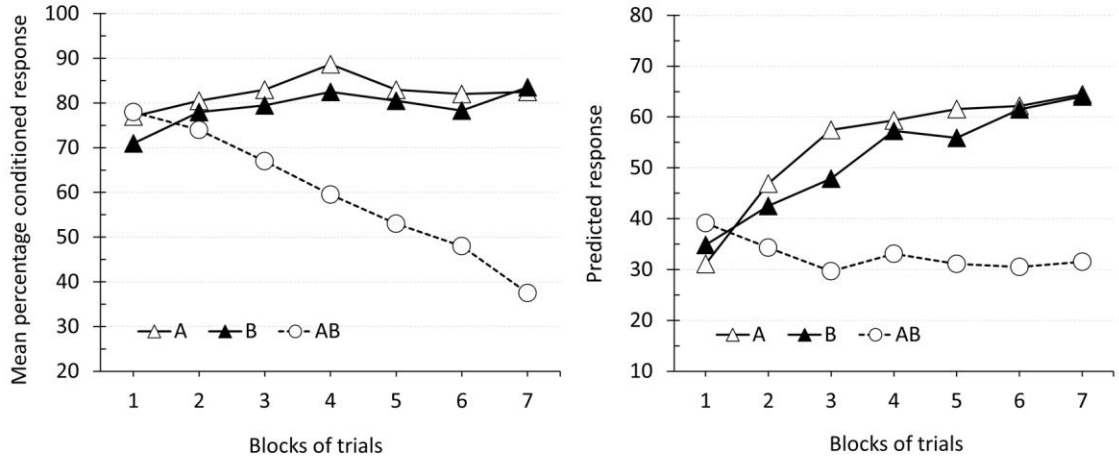


Figure 16: Empirical (original measurement units) and simulated results of Experiment 1 in Whitlow and Wagner (1972). The left panel is an adaptation of the published data showing responses to A, B, and AB during a negative patterning discrimination. The right panel displays the corresponding simulated responses.

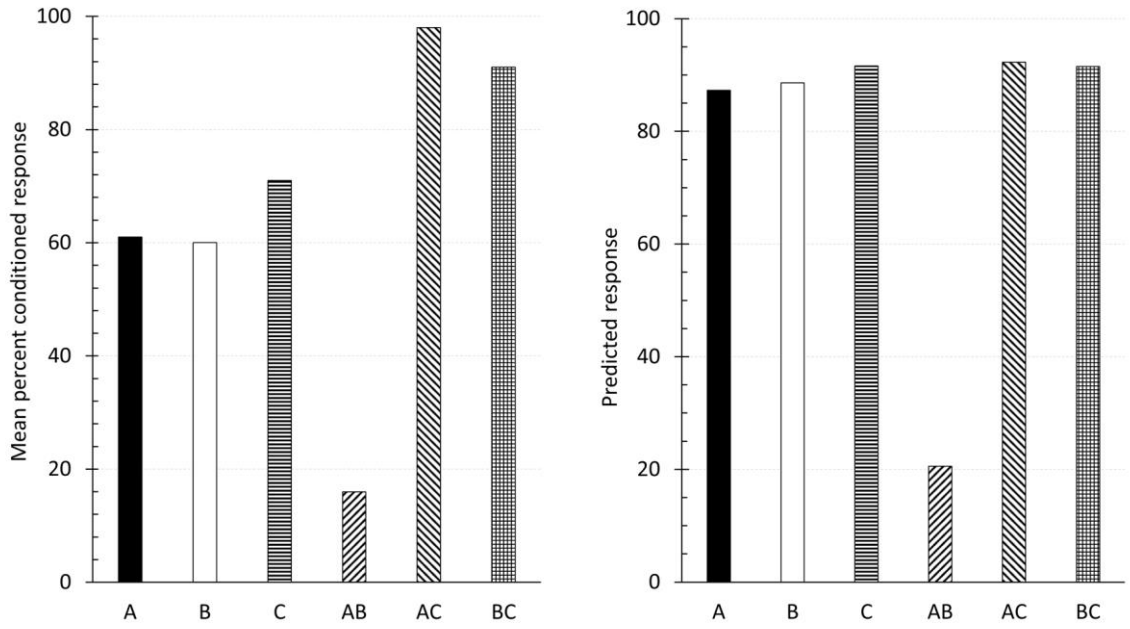


Figure 17: Empirical (original measurement units) and simulated results of Experiment 1 in Whitlow and Wagner (1972). The left panel is an adaptation of the published data showing test data following a negative patterning discrimination. The right panel displays the corresponding simulated responses.

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The temporal parameters for this simulation were a 1 time-unit CS, 1 time-unit US, and 20 time-units ITI. The parameters of the experiment are found in Table 1. Phase 1 consisted of 180 reinforced trials of A, B, and C presented in a random manner. Phase 2 consisted of random 560 presentations of reinforced A and B trials and 1120 non-reinforced AB trials. In Phase 3, 40 reinforced random presentations of A and B were programmed, together with 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 visibly reproduced the experimental data. Figure 16, right panel, shows the acquisition of the NP discrimination where responding to stimuli A and B progressively increased across trials, while responding to AB steadily declined. The right panel of Figure 17 displays the test results. These parallel the empirical data, that is, suppression of responding to AB was larger than to all other stimuli 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 excitatory toward the US, while elements unique to A or B become highly inhibitory. Moreover, on AB trials the common elements between A and B can only be sampled once. Thus, they contribute relatively less to responding than in 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. First, unique elements of A and B in AB trials strongly predict the common elements, thereby delaying the speed of learning by decreasing their associability. This would contribute to sustain a high the expectation for the reinforcer thus leading to a high negative error term for the unique elements. Secondly, the associability rate to the reinforcer, which is initially sustained due to the inconsistent outcome, is revaluated. The DE model predicts that when training under inconsistent outcomes is prolonged (*e.g.*, A being reinforced and non-reinforced through training) and the error term remains high, animals come to learn that the outcome is inconsistently attained and a mechanism to cease sustained attention is activated. This mechanism would become active sooner for the common elements than for unique elements because these elements are active in more trials, therefore accruing evidence faster for the contingency being inherently random. As a consequence, the associability of the common elements is further reduced, pushing the unique elements to become more inhibitory. The net effect is that in 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.

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

When the stimuli are experienced in isolation, the contribution of the inhibitory strength of the unique elements is reduced by half, yet the same quantity of excitatory strength is provided by the common elements. Hence, generating a higher net reinforcer prediction, and thus a higher rate of responding.

Experiment 9: Biconditional Discrimination

In a biconditional discrimination four compounds of two stimuli are reinforced in a manner such that each individual cue receives both reinforced and non-reinforced training (AB+/CD+/AD-/BD-). 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, enhancing the power of this and other elemental models to predict 'configural' learning without the assumption of configural representations.

To prove it, we simulated the biconditional discrimination experiment in (Lober & Lachnit, 2002), which used an aversive skin-conductance conditioning procedure in human participants. For the biconditional treatment, participants received random presentations of reinforced (AB+/CD+) and non-reinforced (AD-/BD-) compounds. The participants learned the discrimination, progressively increasing responding on the reinforced trials and withholding responding on non-reinforced trials (Figure 18, left panel).

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

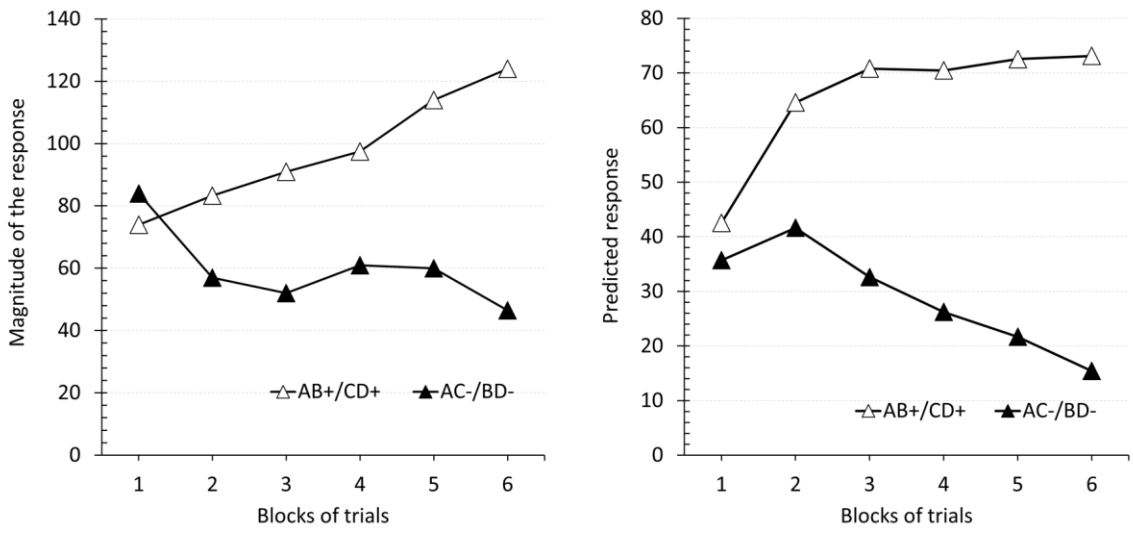


Figure 18: Empirical (original measurement units) and simulated results of a biconditional experiment in Lober and Lachnit (2002). The left panel is an adaptation of the published data showing second-interval responses (SIR) for combined AB+/CD+ trials and AD-/BC-. The right panel displays the corresponding simulated responses.

The simulation's temporal parameters consisted of a one time-unit CS and US, and a 50 time-units ITI. The remaining parameter values are presented in Table 1. Training consisted of 50 randomly presented trials of each reinforced compounds AB+ and CD+, and of the non-reinforced compounds AC- and BD-. The full design is in Table 2.

The simulation matched the empirical pattern of responding. Figure 18, right panel, shows 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 remaining high.

The DE model predicts that initially, the context and the unique elements of the four compound stimuli would gain some excitatory strength. In parallel, the elements common to the stimuli would also gain modest excitatory strength. However, as training progresses the elements common to the two pairs of reinforced trials (*i.e.*, the elements common to A and B and to C and D) would progressively become inhibitory, promoting super-excitation of their unique elements. On reinforced trials, *e.g.*, AB, both A and B would recall their mutual common elements, but they could only be sampled once and therefore their activation would be comparatively lower than that of the unique elements and the US. Thus, due to the

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

dynamic asymptote mechanism, the discrepancy in the level of activation would preclude the common elements from gaining excitatory strength. On the next non-reinforced trial, the prediction error would turn negative (as the US is not present but there is a general excitatory prediction), and all elements would lose strength, but because these common elements would have weaker strengths than other elements, they would reach negative values earlier. At that point, they would foster acquisition of the unique elements excitatory strength on subsequent reinforced trials (by increasing the prediction error). On the next non-reinforced trial, these boosted excitatory elements would produce a larger negative prediction error, engendering more inhibition in the common elements. This cycle would recur, adjusting the strengths through training.

Given that in each reinforced trial (AB and CD) their respective inhibitory common elements are only sampled once the inhibitory impact would be less significant than in the non-reinforced trials (*e.g.*, AC) in which A would retrieve the elements common to AB and C would retrieve the inhibitory elements common to CD. Thus, duplicating the contribution of the inhibitory strength.

Experiment 10: Generalization Decrement Overshadowing vs. External Inhibition

A significant hurdle for configural models of learning is the observation that removing a cue from a previously reinforced compound stimulus (overshadowing test) produces a greater decrement in learning than when a novel cue is added to the compound (external inhibition test). As an elemental model, the DE model inherits an advantage in explaining this effect: training a stimulus compound, *e.g.*, AB, and then testing a single stimulus (*e.g.*, A) should lead to a strong generalization decrement because the summation rule would imply that the total associative strength of the compound is distributed in the individual stimulus strength. Thus, testing a single stimulus removes one source of predictive value and necessarily reduces the total prediction. In contrast, training a single stimulus A implies that the total amount of available prediction is accrued by A. Consequently, testing a compound AB should result in no decrement, provided that B has no inhibitory value itself.

Precisely this difference between external inhibition and overshadowing was studied in Brandon *et al.* (2000), using a rabbit eyeblink conditioning procedure. In their experiment, Group A, Group AB, and Group 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 (see Table 2). The experiment found that adding

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

a cue to a previously reinforced compound produced a smaller decrement in responding than removing a cue from a previously reinforced compound. As it is apparent in the left panel of Figure 19, the decrement of generalization produced by adding one or two cues (AB and ABC) in Group A was comparatively smaller, although clearly present, than that observed when B was removed in Group AB, and when one or two cues were omitted in Group ABC.

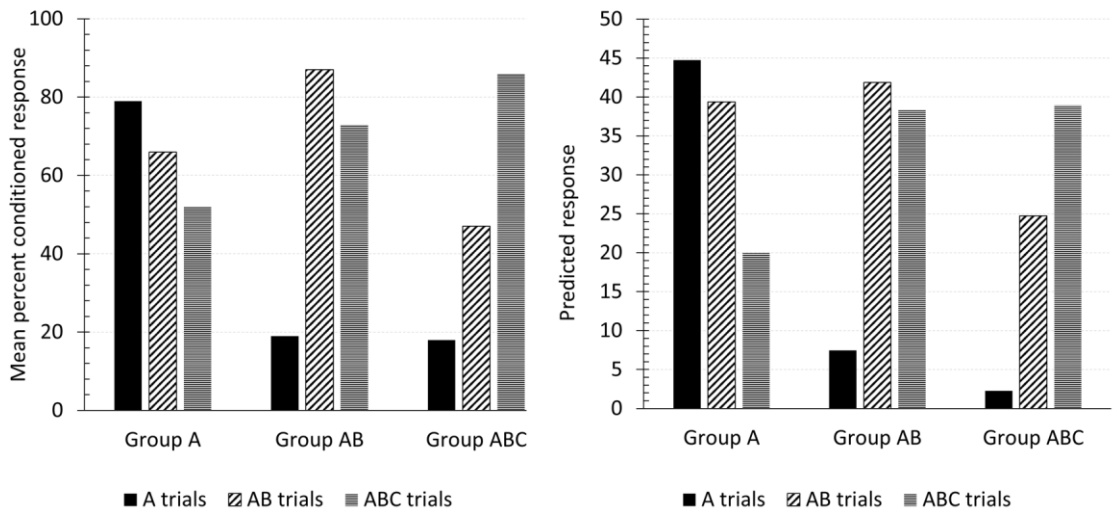


Figure 19: Empirical (original measurement units) and simulated results of a generalization decrement experiment in Brandon *et al.* (2000). The left panel is an adaptation of the published data showing responses during test to A, AB, and ABC. The right panel displays the corresponding simulated responses.

The experiment was simulated using the following temporal parameters: a 5 time-units CS, 1 time-unit US, and 50 time-units ITI. In Phase 1 groups A, AB, and ABC received 20 programmed reinforced trials of type A, AB, and ABC, respectively. Phase 2 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.

Simulated results are an accurate reproduction of the empirical pattern. That is, the generalization decrement adjusted to the incremental change from the conditioned cue to the testing cue in testing, and this decrement in generalization was larger when the cues were removed than when they were added (Figure 19, right panel).

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The DE model produces such a result due to its elemental ontology leading to the summation assumption of constituent associative strengths of a compound. The effect observed by the addition of extra cues can be ascribed to the presence of common elements. Since the test was conducted including reinforced trials, the unique features of the stimuli added could have undergone some inhibitory learning, thus reducing the net compound strength.

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

General Discussion

In this paper, we have introduced a real-time computational model of associative learning, the Double Error (DE) model. Its ontology is founded upon a connectionist network, wherein nodes representing the attributes of any stimulus (elements), whether reinforced or not, enter into association with one another in proportion to their level of activity and are distributed as subsets of shared and unique elements to the stimuli involved. The defining algorithmic characteristics of the DE model are threefold. First, it extends traditional error correction learning, familiar from models such as RW, TD, and SOP, by modulating learning by a predictor error term. This new error term measures the extent to which a predictor of an outcome is in return itself predicted by other stimuli (*e.g.*, the context). An immediate implication is that more predictable predictors associate at a slower rate than less predictable ones. Secondly, the model conceptualizes a revaluation associability that adjusts independently for motivationally relevant and neutral outcomes. Thirdly, the model introduces a novel variable asymptote of learning that encapsulates more closely the principle of Hebbian learning, by postulating that elements with similar activity levels form stronger associations with one another. Hence the direction of learning between two cues at a given time can differ, in an otherwise equivalent condition, based on whether the prior link strength from one to the other is above or below the momentary level of this asymptote. This process critically allows the DE model to predict seemingly contradictory mediated learning effects.

The associability of a cue operates by keeping track over a window of time, through a moving average, of uncertainty in the occurrence of a given category of outcome, reinforcement or otherwise, and thence modulating its respective rate of association. Thus, for instance, when the occurrence of reinforcement following a stimulus is uncertain, learning of associations between the latter and the reinforcer proceeds more rapidly. However, the stimulus associability is revaluated in the face of persistent uncertainty and a decay mechanism is triggered under such conditions. This revaluation in the cue associability 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 that reduces the associability progressively. Though the Mackintosh model's attentional rule seems opposed to that used in the DE model (as the latter resembles the PH alpha), the DE model revaluated alphas can in fact produce effects whereby the best predictor of an outcome captures the most selective attention, due to it producing the highest US prediction error before the onset of US.

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

This, however, is dependent upon sufficiently long CSs being used. Similarly to the McLaren-Mackintosh, SLGK, and SOP models, learning between motivationally neutral cues allows the DE model to reproduce various 'silent learning' and preexposure 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 preexposure. Thus, the model does not require arbitrary memory states for producing an account of stimulus exposure and it postulates the same learning rules and processes both for reinforced and non-reinforced learning, thereby maintaining parsimony. That is, only one learning rule and one type of element are used in the model, *i.e.*, we maintain that reinforced outcomes should be treated simply as a subset of normal stimuli, as opposed to having a completely unique ontology in terms of their stimulus representation and learning. The intensity of a conditioned response to a motivationally relevant stimulus (or to a neutral one, *e.g.*, an orienting response to a light when predicted by a tone) is simply a function of its associative strength. The model's integration of its unique alpha conceptualization together with its distinctive second error term results in it accounting for preexposure effects as arising from the expectancy of a cue decreasing due to learning. It hence avoids difficulties faced by other models, such as those derived from proposing that a noUS link forms during preexposure; a suggestion contradicted by evidence denoting US-specificity of inhibition. The model predicts both context specificity and the sigmoidal shape of latent inhibition, with the latter resulting from the exponential changes in learning speed produced by the revaluation associability. It 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. Furthermore, exposing the context by itself after CS exposure should likewise diminish the latent inhibition effect, as this would weaken the context-CS prediction during subsequent acquisition. In a similar manner, the model reproduces the effect observed by Leung *et. al*. whereby latent inhibition is enhanced by a preexposure of a CS compound. As this effect is explained in the model by the cue that undergoes reinforcement being predicted by the associative representation of the cue it retrieves, the implication is that presenting this retrieved cue in isolation before the acquisition phase should attenuate the observed effect. These preexposure related learning mechanisms also carry over to offering an account of the Hall-Pearce effect, by maintaining that the reinforcement of a CS, using a weak reinforcer, acts in many ways as a form of preexposure treatment from the point of view of the model, due to the lower intensity of the US in the first phase treatment. Therefore, the model's explanation of this phenomena deviates from more

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 to the reinforcer is influenced more by the weak intensity of the US than by CS-US learning. This explanation for the Hall-Pearce effect additionally implies that conducting the second phase of the procedure in a novel context should lessen the observed effect. Lastly, for the effect of perceptual learning, differential CS-CS learning and variable attention shifts between intermixed and blocked presentations of CS preexposure trials allow the model to account for the effect being more pronounced with intermixed CS compounds. Specifically, weaker within-compound associations lead to differential rates of mediated conditioning in later stages of the procedure.

The DE model's dynamic asymptote of learning produces either excitatory or inhibitory learning depending on the causal connection between stimuli as tracked by the distance between the activations of the predicting and predicted stimulus. Through the interplay of this asymptote and its mechanism of associative retrieval the model accounts for learned irrelevance and mediated conditioning. In the case of retrospective conditioned inhibition and learned irrelevance, the claim is made that the poor correlation between the activity of the US representation and that of the CS representation interacts with the dynamic asymptote to induce the CS to form a slight inhibitory link towards the US through context-mediated revaluation. Accordingly, this explanation differs from the account given by the Mackintosh model, which assumes that learned irrelevance arises as an attentional effect and predicts stronger inhibition following uncorrelated than blocked stimulus presentations. As revaluation conditioned inhibition and learned irrelevance are dependent upon context retrieval of the US, the model predicts that strong context-US learning would strengthen the effects.

Mediated conditioning of a CS is produced in the DE 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 tends to be excitatory and proportional to the degree to which the CS representation is retrieved. Consequently, 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

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

instance, though the asymptote of the retrieved CS in backward blocking and mediated conditioning is the same, the direction of learning in the DE model follows an opposite direction for these phenomena 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 – Dickinson and Burke (1996) and Holland (1983) – or the negative learning rate used in many models for mediated learning, which would each expect mediated learning to follow the same direction for the retrieved CS in both procedures. 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.

Finally, while retaining an elemental nature, the DE model can also reproduce a variety of non-linear discriminations without assuming that an animal has access to *de novo* configural information. It does so through a multi-factorial approach. The fully-connected network architecture of the model, 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. 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 cannot be sampled twice at the same time). The DE model's variable revaluation alpha can amplify discrimination learning through a unique decay process, which lowers the associability of cues presented with a persistently uncertain outcome. Silent learning and attentional processes, as mentioned before, produce unitization of common and unique elements, as well as latent inhibition and a greater 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. The model also 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 non-linear discrimination effects, including negative patterning, biconditional discriminations, and perceptual learning using a completely elemental representation.

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

Ahveninen, J., Kähkönen, S., Tiitinen, H., Pekkonen, E., Huttunen, J., Kaakkola, S., Ilmoniemi, R. J., & 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. & Mondragón, E. (2014). Computational Models of Classical Conditioning - A Qualitative Evaluation and Comparison. In B. Duval, J. van den Herik, S. Loiseau & J. Filipe (Eds.), *Proceedings of the 6th International Conference on Agents and Artificial Intelligence* (pp. 544–547). Setúbal, Portugal: SCITEPRESS.

Alonso, E., Mondragón, E., & Fernández, A. (2012). A Java simulator of Rescorla and Wagner's prediction error model and configural cue extensions. *Computer Methods and Programs in Biomedicine*, *108*(1), 346–355.

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

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

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

Balkenius, C., & Morén, J. (1998). Computational models of classical conditioning: A comparative study. In R. Pfeifer, B. Blumberg, J.-A. Meyer & S. W. Wilson (Eds.), *From Animals to Animats 5* (pp. 348-353). Cambridge, MA: MIT Press.

Blair, C. A. J., & 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.

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., & 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.

Brandon, S. E., Vogel, E. H., & 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., & 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.

Bush, R. R., & Mosteller, F. (1955). Stochastic models for learning. John Wiley & Sons, Inc.

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

Colwill, R. M., & 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., Aitken, M. R. F., Shanks, D. R., Robbins, T. W., Bullmore, E. T., Dickinson, A., & Fletcher, P. C. (2007). Disrupted prediction-error signal in psychosis: evidence for an associative account of delusions. *Brain*, *130*(9), 2387-2400.

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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., &Burke, J. (1996). Within-compound associations mediate the retrospective revaluation of causality judgements. *Quarterly Journal of Experimental Psychology*, *37B*, 397–416.

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

Dwyer, D. M. (1999). Retrospective revaluation or mediated conditioning? The effect of different reinforcers. *The Quarterly Journal of Experimental Psychology*, *52*(4), 289-306.

Gallistel, C. R., Fairhurst, S., & 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.

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

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

Ginsburg, S., & 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., & 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., & Nef, P. (2001). Ca2+ Signaling via the Neuronal Calcium Sensor-1 Regulates Associative Learning and Memory in C. elegans. *Neuron*, *30*(1), 241–248.

Hall, G. (1991). *Perceptual and associative learning.* Oxford, UK: Clarendon Press-Oxford University Press.

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

Hall, G., & 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., & 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., & Rodriguez, G. (2010). Associative and nonassociative processes in latent inhibition: An elaboration of the Pearce-Hall model. In R.E. Lubow & I. Weiner (Eds.), *Latent inhibition: Cognition, neuroscience, and applications to schizophrenia* (pp. 114-136). Cambridge, England: Cambridge University Press.

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

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

Haselgrove, M., & Hogarth, L. (2012). *Clinical Applications of Learning Theory*. Hove, England: Psychology Press.

Holland, P. C. (1983). Representation-mediated overshadowing and potentiation of conditioned aversions.

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Journal of Experimental Psychology: Animal Behavior Processes, *9*(1), 1–13.

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

Holland, P. C., & 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*. Oxford, England: Appleton-Century.

Jie, H. L. (2008). *Neuroimaging of associative learning* (master's thesis). National University of Singapore, Singapore.

Kamin, L. J. (1968). "Attention-like" processes in classical conditioning. In M.R. Jones (Ed.), *Miami Symposium on the Prediction of Behavior, 1967: Aversive Stimulation* (pp. 9-31). Coral Gables, FL: University of Miami Press.

Kamin, L. J. (1969). Selective association and conditioning. In N.J. Mackintosh & W.K. Honig (Eds.), *Proceedings of the Symposium on Fundamental issues in Associative Learning* (pp. 42-64). Halifax, Canada: Dalhousie University Press.

Kehoe, E. J., Schreurs, B. G., & 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., & Poo, M. (2004). Spike Train Timing-Dependent Associative Modification of Hippocampal CA3 Recurrent Synapses by Mossy Fibers. *Neuron*, *41*(3), 445–454.

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

Kruschke, J. K. (2011). Models of attentional learning. In: E. M. Pothos & A. J. Wills (Eds.), *Formal Approaches in Categorization* (pp. 120-152). Cambridge, England: Cambridge University Press

G., & 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., & 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.

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., Haselgrove, M., & Esber, G. R. (2012). Modeling attention in associative learning: Two processes or one? *Learning & Behavior*, *40*(3), 292–304.

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

Leung, H. T., Killcross, A. S., & 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., & Lachnit, H. (2002). Configural learning in human Pavlovian conditioning: acquisition of a biconditional discrimination. *Biological Psychology*, *59*(2), 163–168.

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

Ludvig, E. A., Bellemare, M. G., & Pearson, K. G. (2011). A primer on reinforcement learning in the brain: Psychological, computational, and neural perspectives. In E. Alonso & E. Mondragón, *Computational Neuroscience for Advancing Artificial Intelligence: Models, Methods and Applications* (pp. 111-144). Hershey, PA: IGI Global.

Ludvig, E. A., & Koop, A. (2008). Learning to Generalize through Predictive Representations: A Computational Model of Mediated Conditioning. In M. Asada, J. C. T. Hallam, J.-A. Meyer & J. Tani, *From Animals to Animats 10* (pp. 342–351). Berlin, Heidelberg: Springer Berlin Heidelberg.

Ludvig, E. A., Mirian, M. S., Kehoe, E. J., & Sutton, R. S. (2017). Associative Learning from Replayed Experience. *bioRxiv Preprint*. doi: https://doi.org/10.1101/100800

Ludvig, E. A., Sutton, R. S., Verbeek, E., & Kehoe, E. J. (2009). A computational model of hippocampal function in trace conditioning. In D. Koller, D. Schuurmans, Y. Bengio & L. Bottou (Eds.), *Advances in Neural Information Processing Systems 21* (pp. 993–1000). Red Hook, NY: Curran Associates, Inc.

Mackintosh, N. J. (1973). Stimulus selection: Learning to ignore stimuli that predict no change in reinforcement. In R. A. Hinde & J. Stevenson-Hinde, *Constraints on learning: Limitations and predispositions* (pp. 75-96). Oxford, England: Academic Press.

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., Kaye, H., & Bennett, C. H. (1991). Perceptual learning in flavour aversion conditioning. *The Quarterly Journal of Experimental Psychology*, *43*(3), 297–322.

Marschner, A., Kalisch, R., Vervliet, B., Vansteenwegen, D., & Büchel, C. (2011). Neural correlates of human associative learning. *Tsinghua Science & Technology*, *16*(2), 140–144.

McLaren, I. P. L. (1993). APECS: A solution to the sequential learning problem. In M. Ringle (Ed.), *Proceedings of the Fifteenth Annual Convention of the Cognitive Science Society* (pp. 717–722). Hillsdale, NJ: Lawrence Erlbaum.

McLaren, I. P. L., & 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. R., Barnet, R. C., & Grahame, N. J. (1995). Assessment of the Rescorla-Wagner model. *Psychological Bulletin*, *117*(3), 363-386.

Miller, R. R., & 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., & Witnauer, J. E. (2016). Retrospective revaluation: The phenomenon and its theoretical implications. *Behavioural Processes*, *123*, 15–25.

Mondragón, E., Alonso, E., Fernández, A., & Gray, J. (2013). An extension of the Rescorla and Wagner Simulator for context conditioning. *Computer Methods and Programs in Biomedicine*, *110*(2), 226–230.

Mondragón, E., Gray, J., & Alonso, E. (2013). A Complete Serial Compound Temporal Difference Simulator for Compound stimuli, Configural cues and Context representation. *Neuroinformatics*, *11*(2), 259–261.

Mondragón, E., Gray, J., Alonso, E., Bonardi, C., & Jennings, D. J. (2014). SSCC TD: A Serial and Simultaneous Configural-Cue Compound Stimuli Representation for Temporal Difference Learning. *PLoS*

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ONE, *9*(7), e102469.

Mondragón, E., & 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., & 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.

Montague, P. R., Dayan, P., & 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., & Brunzell, D. H. (1998). Predictive timing under temporal uncertainty: The time derivative model of the conditioned response. In D. A. Rosenbaum & C. E. Collyer (Eds.), *Timing of behavior: Neural, psychological, and computational perspectives* (pp. 3-34). Cambridge, MA: The MIT Press.

Murphy, R.A., Mondragón, E. & Murphy, V. A. (2008). Rule learning by rats.*Science*, *319*(5871), 1849-1851.

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., & O'Doherty, J. P. (2012). Neural Prediction Errors Reveal a Risk-Sensitive Reinforcement-Learning Process in the Human Brain. *Journal of Neuroscience*, *32*(2).

Panayi, M. C., & 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, England: Oxford University Press.

Pearce, J. M. (1987). A model for stimulus generalization in Pavlovian conditioning. *Psychological Review*, *94*(1), 61-73.

Pearce, J. M., & Bouton, M. E. (2001). Theories of associative learning in animals. *Annual Review of Psychology*, *52*, 111-39.

Pearce, J. M., George, D. N., & 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., & 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., & 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., & 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.

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. (1970). Reduction in the effectiveness of reinforcement after prior excitatory conditioning. *Learning and Motivation*, *1*(4), 372–381.

Rescorla, R. A. (1971a). Variation in the effectiveness of reinforcement and nonreinforcement following prior inhibitory conditioning. *Learning and Motivation*, *2*(2), 113-123.

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Rescorla, R. A. (1971b). Summation and retardation tests of latent inhibition. *Journal of Comparative and Physiological Psychology*, 75(1), 77-81.

Rescorla, R. A. (1972). "Configural" conditioning in discrete-trial bar pressing. *Journal of Comparative and Physiological Psychology, 79*(2), 307-317.

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., & Durlach, P. J. (1985). Analysis of the unique cue in configural discriminations. *Journal of Experimental Psychology: Animal Behavior Processes, 11*(3), 356-366.

Rescorla, R. A., & Wagner, A. R. (1972). *A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement.* In A. H. Black & W. F. Prokasy (Eds.). *Classical Conditioning II: Current Research and Theory* (pp. 64-99). New York: Appleton Century Crofts.

Roesch, M. R., Esber, G. R., Li, J., Daw, N. D., & Schoenbaum, G. (2012). Surprise! Neural correlates of Pearce-Hall and Rescorla-Wagner coexist within the brain. *European Journal of Neuroscience*, *35*(7), 1190–1200.

Saavedra, M. A. (1975). Pavlovian compound conditioning in the rabbit. *Learning and Motivation*, *6*(3), 314–326.

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

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

Schultheis, H., Thorwart, A., & 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), 139-47.

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 6:24*.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science 275*(14), 1593-9.

Sutton, R. S., & Barto, A. G. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review*, *88*(2), 135-70.

Sutton, R. S., & Barto, A. G. (1987). A temporal-difference model of classical conditioning. In J. D. Moore & J. F. Lehman (Eds.), *Proceedings of the Ninth Annual Conference of the Cognitive Science Society* (pp. 355–378). Mahwah, NJ: Erlbaum.

Symonds, M., & 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.

Turkkan, J. S. (1989). Classical conditioning beyond the reflex: An uneasy rebirth. *Behavioral and Brain Sciences*, *12*(1), 161–179.

Urushihara, K., & Miller, R. R. (2010). Backward blocking in first-order conditioning. *Journal of Experimental*

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Psychology. Animal Behavior Processes, *36*(2), 281–95.

Wagner, A.R. (1981). SOP: A model of automatic memory processing in animal behavior. In N.E. Spear & R.R. Miller (Eds.), *Information processing in animals: Memory mechanisms* (pp. 5–47). Hillsdale, NJ: Erlbaum.

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., & Brandon, S. E., (2001) A componential theory of Pavlovian Conditioning. In R.R. Mowrer and S.B. Klien (Eds.), *Handbook of Contemporary Learning Theories* (pp. 23–64). Mahwah, NJ: Erlbaum.

Wagner, A. R., & Rescorla, R. A. (1972). Inhibition in Pavlovian conditioning: Application of a theory. In R.A. Boakes & M.S. Halliday (Eds.), *Inhibition and learning* (pp. 301–336). London, England: Academic Press.

Ward-Robinson, J., & Hall, G. (1996). Backward sensory preconditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, *22*(4), 395.

Ward-Robinson, J., & Hall, G. (1998). Backward Sensory Preconditioning When Reinforcement is Delayed. *The Quarterly Journal Of Experimental Psychology*, *51B*(4), 349–362.

Whitlow, J. W., & Wagner, A. R. (1972). Negative patterning in classical conditioning: Summation of response tendencies to isolable and configurai components. *Psychonomic Science*, *27*(5), 299–301.

Wilson, P. N., & 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., & Preston, A. R. (2012). Hippocampal and Ventral Medial Prefrontal Activation during Retrieval-Mediated Learning Supports Novel Inference. *Neuron*, *75*(1), 168–179.

Table 1: Simulation parameters

Fixed Model Parameters

Element Representation	CV	20
	Curve right skew	20
	Set size	10
	Common elements proportion	0.4
Activation discount	CV CS element activation	2
	Associative activated discount	0.75
Memory Discounts	Backward discount	0.75
	Eligibility Trace discount	0.95
Model	Associability Recency	0.1

Variable Parameters

			Exp. 1	**Exp. 2**	**Exp.3**	**Exp. 4**	**Exp. 5**	**Exp. 6**	**Exp. 7a**	**Exp. 7b**	**Exp. 8**	**Exp. 9**	**Exp. 10**
Associability (α)	αr	CS	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80
		Ctx	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
	αn	CS	0.40	0.80	0.40	0.80	0.40	0.80	0.80	0.80	0.40	0.40	0.80
		Ctx	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20
	α+	US	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20
Salience	s	CS	0.10	0.35	0.30	0.05	0.50	0.15	0.80	0.50	0.30	0.10	0.03
		Ctx	0.10	0.20	0.07	0.01	0.10	0.015	0.07	0.07	0.07	0.07	0.01
	β	US	0.15	0.90	0.75	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90

Table 2: All experimental designs that are to be simulated

n	Experiment	Group / *Condition*	Phase 1	Phase 2	Phase 3	Phase 4
1	Latent Inhibition and Context Specificity. *Hall & Channel, 1983*	**Exposed Same**	φ : A-	φ:A+		
		Exposed Different	ψ: A+	φ:A+		
		Control Same	φ	φ:A+		
		Control Different	ψ	φ:A+		
2	Compound latent inhibition. *Leung, Killcross, & Westbrook, 2011*	**Element**	A-/B-/C-	AB-/C-	C+	C-
		Compound	A-/B-/C-	AB-/C-	A+	A-
3	Hall-Pearce Effect. *Hall & Pearce, 1979*	**Tone-shock**	T+	T++		
		Light-shock	L+	T++		
		Tone-alone	T-	T++		
4	Learned Irrelevance and Retrospective Correlation. *Baker, Murphy & Mehta, 2003*	**N/Sh (1A)**		N-/+/N+	N+	
		N+Sh (1A)		N- +	N+	
		N/Sh (1B)	L+	N-/+/N+	L-/NL-	L+/NL-
		N/Sh (1B)	L+	N- +	L-/NL-	L+/NL-
5	Perceptual Learning. *Blair & Hall, 20013*		AX-/BX- & CX-	AX+	BX-/CX-	
6	Mediated Learning. *LePelley & McLaren, 2001*	*A2-A2*	AB-/C+	AC-	A/B/C/D/E/F/G/H/I/J/K /L/M/N/O/P/Q/R	
		A2-A1	DE-/F-	DF+		
		Control	GH-/I+	GJ-		
		Unovershadowing	KL+	K-		
		B. Block	MN+	M+		
		RR Control	OP+	O+/O-		
		Fillers	Q-	QR+		
7a	Backward Sensory Preconditioning. *Ward-Robinson & Hall (1996)*	**Ext**	A→X/B→Y	A→US/B	A-/B-	X-/Y-
		VI	A→X/B→Y	A→US/B		X-/Y-
7b	Mediated Extinction. *Ward-Robinson & Hall (1996)*		A→X/B→Y	X+/Y+	A-	X-/Y-
8	Negative Patterning. *Whitlow & Wagner (1972)*		A+/B+/C+	A+/B+/AB-	A+/B+/AB-/C+	A-/B-/C-/AB-/AC-/BC-
9	Biconditional Discrimination. *Lober & Lachnit, 2002*		AB+/CD+/AC-/BD-			
10	External inhibition vs. overshadowing. *Brandon, Vogel, & Wagner, 2000*	**A**	A+	A+/A-/AB+/AB-/ABC+/ABC-		
		AB	AB+	A+/A-/AB+/AB-/ABC+/ABC-		
		ABC	ABC+	A+/A-/AB+/AB-/ABC+/ABC-		

bioRxiv preprint doi: https://doi.org/10.1101/210674; this version posted October 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.