©American Psychological Association, 2020. This paper is not the copy of record and may not exactly replicate the authoritative document published in the APA journal. Please do not copy or cite without author's permission. The final article is available, upon publication, at: https://doi.org/10.1037/xan0000252.

The representation of stimulus conjunction in theories of associative learning: A context-dependent added-elements model

David N. George

Department of Psychology, University of Hull, Hull, UK

Address for correspondence:
David N. George
Department of Psychology
University of Hull
Hull HU6 7RX
Tel: +44 (0) 1482 465483
Email: d.george@hull.ac.uk

© 2020, American Psychological Association. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors' permission. The final article will be available, upon publication, via its DOI: 10.1037/xan0000252

Abstract

This paper briefly reviews three theories concerning elemental and configural approaches to stimulus representation in associative learning and presents a new Context-dependent Added Elements Model (C-AEM). This model takes an elemental approach to stimulus representation where individual stimuli are represented by single units and stimulus compounds activate both those units and configurational units corresponding to each conjunction of two or more stimuli. Activity across these units is scaled such that each stimulus always contributes the same amount of activity to the system whether they are presented in isolation or in compound; the configurational units 'borrow' activity from representation units for individual stimuli (and from each other). This scaling is affected by the extent to which stimuli interact with each other perceptually. Hence, the model is conceptually similar to Wagner's (2003) Replaced Elements Model but lacks features that explicitly code for the absence of stimuli (i.e., inhibited elements). Simulations of the model are reported for a range of generalization and discrimination learning tasks, conflicting results from which have previously been taken to provide support for either configural or elemental theories of learning.

Keywords: Pavlovian conditioning, configural, elemental, generalization, discrimination learning

The representation of stimulus conjunction in theories of associative learning: A context-dependent added-elements model

Pavlov (1927) reported that if a 1 kHz tone was paired with food such that it came to provoke a conditioned response (CR), tones of other frequencies would also trigger the CR. Furthermore, the strength of the response diminished as the difference in the frequencies of the training and test stimuli increased. Models of associative learning have traditionally explained this generalization of responding from one stimulus to another by appealing to similarities in the internal representations of patterns of stimulation. One approach, exemplified by Spence's (1936, 1937) theory of discrimination learning and by stimulus sampling theory (Atkinson & Estes, 1963; Estes 1950, 1955a, 1955b) is to assume that the representation of any stimulus situation consists of a number of theoretic elements, or micro-features, each of which may become associated with a response. The extent to which responding will generalize from one situation to another is determined by the degree of overlap between the populations of elements that represent these different stimulus situations. Such *elemental* representational schemes have proved popular and form the basis of a number of contemporary models of learning (e.g., Harris, 2006; McLaren & Mackintosh, 2000, 2002).

Despite the success of elemental models of learning, an alternative, *configural*, approach to stimulus generalization was proposed by Pearce (1987, 1994, 2002) in response to the results of a number of experiments examining the influence of similarity on animals' ability to learn to discriminate between different patterns of stimulation. Whereas elemental models assume that individual stimulus elements enter into direct associations with the representation of the outcome (unconditioned stimulus, US), Pearce's configural theory supposes that any pattern is represented as a whole, and it is this *configural representation* that enters into an association with the US. Generalization occurs within this model because one pattern may partially activate the representation of another pattern as a function of their similarity.

Other models represent stimuli elementally, with those representations entering into direct associations with the US, but they make the additional assumption that when two or more stimuli are presented in compound a unique configurational cue is generated. This configurational cue may itself enter into associations (Spence, 1952; Wagner & Rescorla, 1972; Brandon, Vogel & Wagner, 2000; Wagner, 2003).

In this article I shall first briefly describe some features of three models that were, over many years, a focus of debate between the laboratories of Pearce and Wagner: the Rescorla-Wagner (RW) elemental model (Rescorla & Wagner, 1972), Pearce's (1987, 1994, 2002) Configural Theory and Wagner's (2003) Replaced Elements Model (REM). I shall then describe a new elemental model of learning, the Context-dependent Added Elements Model (C-AEM), which is based on the RW model, and is computationally much less demanding than REM. Finally, I will review some of the key experimental results that provide differential support for elemental and configural theories and explore whether C-AEM can reconcile conflicting findings from Pearce's and Wagner's laboratories.

The Rescorla-Wagner Model

According to Rescorla & Wagner (1972), the magnitude of the CR provoked by a conditioned stimulus (CS) is determined by the strength of an associative link between internal representations of the CS and a US. When the CS is paired with a US the strength of this association (the associative strength of the CS) is updated, with the change in the associative strength being determined by the extent to which the US is surprising, or unexpected. Equation 1 shows how the change in the associative strength, *V*, of stimulus A is determined by the sum of associative strengths of all stimuli present on the conditioning trial, ΣV, the maximum associative strength supported by the US, λ, and learning rate parameters, α_A and β, associated with the CS and the US, respectively. λ is typically assigned some arbitrary positive value (frequently 1) when the US is present and is set to zero when the US is absent. The learning rate parameters take some value between zero and 1.

$$\Delta V_A = \alpha_A \beta \left(\lambda - \Sigma V \right) \quad (1)$$

A consequence of conditioning with a stimulus compound is that each individual stimulus may accrue associative strength. It follows from Equation 1, however, that following simple conditioning the asymptotic associative strength of any element of a compound is determined by two factors: the number of other stimuli in the compound, and the relative values of the learning rate parameter, α, associated with each stimulus. The asymptotic associative strength of stimulus A when paired with a US in compound with other CSs is given by Equation 2.

$$V_A = \frac{\alpha_A}{\Sigma \alpha} \lambda \quad (2)$$

The use of the combined error term (λ – Σ*V*) in Equation 1 sets the RW model apart from the linear operator rule of Bush & Mosteller (1955) and allows the RW model to account for learning phenomena such as cue competition effects (e.g., blocking, overshadowing, over-expectation, super-conditioning) and the development of conditioned inhibition. There are, however, certain discrimination learning tasks which pose considerable problems for the RW model. These include patterning and conditional discrimination tasks which have no linear solution. Saavedra (1975) trained rabbits on a biconditional discrimination involving four compound stimuli. Compounds AC and BD were each paired with an intra-orbital electric shock, whereas compounds AD and BC were presented in the absence of this US. Because each of the individual cues, A, B, C, and D, occurred equally often on reinforced and non-reinforced trials, the RW model, as characterized above, would predict that each cue should gain a moderate amount of associative strength and all four compounds should provoke a CR of similar magnitude. That Saavedra's rabbits came to respond in the presence of AC and BD but learnt to withhold responding to AD and BC cannot be explained by this purely elemental model of learning.

One solution to this problem is to assume that the relationships between components of specific combinations of stimuli may give rise to unique *configurational cues* (e.g., Spence, 1952;

Wagner, 1971; Wagner & Rescorla, 1972). Hence, we may recast Saavedra's discrimination task as AC*w*+ BD*x*+ AD*y*ø BC*z*ø, where + and ø denote the presentation or omission of the US, and *w*, *x*, *y*, and *z* are the configurational cues generated by the compounds AC, BD, AD, and BC, respectively. If we allow these configurational cues to gain associative strength in the same way as normal stimuli, then the RW model can solve the biconditional discrimination with the stimuli A, B, C, and D acquiring relatively little associative strength. The configurational cues *w* and *x*, which are generated by the reinforced patterns AC and BD, are predicted to become moderately excitatory, whereas *y* and *z* which are generated by the non-reinforced patterns AD and BC should become moderately inhibitory[1].

Pearce's Configural Theory

Pearce (1987, 1994, 2002) proposed an alternative model of associative learning based on the principle that animals learn about the relationship between entire configurations of stimuli and the outcome. Pearce presented his model within a connectionist framework, consisting of an input network, a layer of configural units, and finally a layer of US units. When a pattern of stimulation is presented to the network, each stimulus within it activates a corresponding unit within the input network and activation of these units is fed forward to the units in the configural layer. The connections within the input network, and those between it and the configural layer, are weighted such that when any pattern– whether it consists of a single stimulus or of several stimuli – is presented to the network it will fully activate a single configural unit which has been recruited to represent that pattern specifically. That configural unit can then enter in an association with a US unit. Following a conditioning trial with pattern j, the associative strength of configural unit j, E_j, will be modified according to a learning rule similar to that employed by the RW model. The increment in associative strength, ΔE_j, is dependent upon learning rate parameters α and β which reflect the conditionability of the configural unit and properties of the US, respectively, and V_j which is the overall level of activation of the US unit when pattern j is presented. The learning rule is shown in Equation 3.

$$\Delta E_j = \alpha\beta(\lambda - V_j) \quad (3)$$

The value of V_j is not necessarily the same as the associative strength of configural unit j, since generalization between patterns means that other configural units may also be partially activated by pattern j, and these units may in turn excite US units. For example, if a network has been exposed to patterns A and AB, then when stimulus A is presented, the configural unit for pattern AB will also be partially activated due to the similarity of pattern AB to A. The extent to which these other, partially activated configural units contribute to the overall activation of the US unit is determined by the level of activation of each configural unit and the strength of its association with the US unit. Accordingly, the activation of the US unit is given by Equation 4. Here, *n* is the total number of configural units in the network, and $S_{j,i}$ is the similarity of pattern j to pattern i.

$$V_j = E_j + \sum_{i=1, i\neq j}^{n} S_{j,i} E_i \quad (4)$$

Finally, the similarity of two patterns i and j is a function of the number of features they have in common N_c, and the total number of features in the two patterns, N_i and N_j. This relationship is shown in Equation 5, which allows similarity to vary in the range 0 to 1.

$$S_{j,i} = \frac{N_c^{\,2}}{N_i N_j} \quad (5)$$

Kinder and Lachnit (2003, see also Pearce, Esber, George & Haselgrove, 2008) suggested that this similarity function may be modified to include a parameter reflecting the discriminability of the individual stimuli that constitute patterns. First, Equation 5 may be re-written as Equation 6. Next, the exponent on the right-hand side of the equation is replaced with the discriminability parameter, *d*. The modified similarity function is given in Equation 7. Increasing the value of the *d* results in a sharpening of gradients of generalization around patterns, and decreasing its value flattens those gradients.

$$S_{j,i} = \left(\frac{N_c}{\sqrt{N_i} \times \sqrt{N_j}} \right)^2 \quad (6)$$

$$S_{j,i} = \left(\frac{N_c}{\sqrt{N_i} \times \sqrt{N_j}} \right)^d \quad (7)$$

Wagner's Replaced Elements Model

There are a number of situations in which the RW model and Configural Theory make contrasting predictions, but where the predictions of each model are supported by empirical evidence. Myers, Vogel, Shin and Wagner (2001, p.41) commented that "Pearce and his associates have consistently found support for predictions of the Pearce (1987) model that challenge the Rescorla and Wagner (1972) model. We have just as consistently obtained results which support the predictions of the Rescorla-Wagner model and challenge the Pearce model." Wagner (2003) attempted to reconcile these different patterns of results by appealing to differences in the stimuli employed in experiments conducted within his own, and Pearce's laboratories. Wagner's experiments that found support for the RW model involved rabbits eyeblink conditioning with stimuli drawn from different modalities, whereas Pearce's experiments that support Configural Theory have involved pigeon autoshaping with only visual cues.

Wagner and Brandon (Brandon & Wagner, 1998; Wagner & Brandon, 2001; Wagner, 2003) suggested that any stimulus is represented as a collection of hypothetical elements, or micro-features. Some elements are context independent and are activated whenever the stimulus is presented. The activation of other elements, however, is determined by the context in which the stimulus is presented. These context dependent elements may be sensitive to the conjunction of certain stimuli (that is, they act as configurational cues), or they may be active only when the stimulus is presented by itself, but not when it is presented in compound with another stimulus (inhibited elements). Hence, if we consider the elements that represent one stimulus (A) that may be presented by itself or in

compound with a single other stimulus (B), then three distinct populations of elements may be identified: context independent elements (A_i) that are activated whenever stimulus A is presented; context-dependent added elements (A_b) that are activated only when A is presented in compound with B; and context-dependent inhibited elements ($A_{\sim b}$) that are activated when A is presented alone and never when it is presented in compound with B. When A is presented in compound with B, activity of the added elements, A_b, can be said to replace the activity of the inhibited elements, $A_{\sim b}$. The proportion of a stimulus' elements that are replaced when it is presented in compound with another is determined by the parameter *r*, and the proportion of elements that are activated whenever the stimulus is presented (i.e., the context independent elements, A_i) is given by $s = (1 - r)$.

Elements are binary; they are either active or inactive, and each may enter into an association with a representation of the US. Changes in the associative strength of active elements are governed by the same rule as for the RW model, given in Equation 1. There are three additional assumptions of the model that are important in determining the precise predictions of REM. First, if stimulus A may be presented in compound with two or more other stimuli (e.g., B and/or C), then replacement of A elements by B and C is statistically independent. That is, some of the elements of A that are context independent with respect to B may be context dependent with respect to C, and vice versa. This principle is captured by equations 8a-8d, presented by Wagner (2003), where Pa is the proportion of elements common to stimulus A and a compound of A with stimuli B, C, ...N. When A is presented in compound with B, a proportion (r_b) of A's representational elements will be inhibited, meaning that the two patterns share $(1 - r_b) = s_b$ elements. Addition of stimulus C to the compound (ABC), will further inhibit a proportion of all A elements active in compound AB (including A_b added elements). Hence, the proportion of A elements common to A and a compound containing A may be determined by multiplying together the *s* values of all of the stimuli presented with A.

$$Pa(A \cap AB) = (1 - r_b) = s_b \tag{8a}$$

$$Pa(A \cap AC) = (1 - r_c) = s_c \tag{8b}$$

$$\supset Pa(A \cap ABC) = (1 - r_b)(1 - r_c) = s_b s_c \tag{8c}$$

$$Pa(A \cap A...N) = \prod_{k=1}^{n-1} s_k \tag{8d}$$

Second, replacement is not random. It is always the same representational elements of A that are replaced in the context of B, always the same elements are activated independently of the context B, and always the same elements that are content-dependently activated in the presence of B.

Third, the degree to which elements of stimulus A are replaced when it is presented in compound with stimulus B (and hence, the parameters r_b and s_b) is determined, to some extent at least, by perceptual properties of A and B. Stimuli which belong to the same modality are assumed to interact with each other at a perceptual and representational level to a greater extent than stimuli which belong to different modalities.

If we accept that replacement is greater for stimuli within the same modality than for stimuli taken from different modalities, REM can account for some of the conflicting results of experiments conducted in Pearce's and Wagner's laboratories (such as simple and differential summation). The same is not the case, however, for complex negative patterning discriminations of the form A+ B+ C+ AB+ AC+ BC+ ABCø. REM is also computationally quite complex. As the number of stimuli which may be presented in combination increases, there is an exponential explosion in the number of ways that they may interact and, consequently, the number of distinct populations of representational elements that must be considered. Because each additional stimulus may interact with existing populations of representation elements in three ways (context independent, context dependently activated, or context dependently inhibited), this expansion in populations may be derived from trinomial theorem (see Appendix A). For any system consisting of n stimuli, $n3^{n-1}$ distinct populations of elements may be

identified. Hence, the number of populations of elements required to model systems in which 1, 2, 3, 4, or 5 stimuli may interact is 1, 6, 27, 108, and 405, respectively.

To address these two issues with REM I present a simplified model of stimulus interaction. The model is inspired by REM but based more closely on the principles of Wagner & Rescorla's (1972) configurational cue model. It has the twin benefits of being capable of explaining a wider range of experimental findings and of being computationally significantly less demanding than REM.

A Context Dependent Added Elements Model

The model incorporates features of Wagner and Rescorla's (1972) configurational cue extension to the RW model and of Wagner's (2003) REM. This Context-dependent Added Elements Model (C-AEM) is based on the assumptions that a) whenever two or more stimuli are presented in compound they generate unique configurational cues, and b) the relative activation of elemental stimulus representations and the configurational cues varies as a function of a parameter, *r*, which reflects the degree to which representations of the stimuli interact with each other. C-AEM departs from REM in two fundamental aspects. First, it does not invoke the notion of inhibited elements and, hence, replacement. Second, it is based on unitary representations of stimuli and of configurational cues. C-AEM is conceptually similar to (but not the same as) a version of REM in which replacement of elements is determined randomly without having to make assumptions about the nature of the population from which a random selection of added elements is sampled.

The representational structure of a stimulus compound in C-AEM is the same as that in the RW model if it is assumed that a configurational cue for each combination of stimuli within that compound is generated. Hence, stimulus A will activate a single unit (*a*); the compound AB will activate units that represent stimuli A and B (*a, b*) and their conjunction (*ab*); the compound ABC will activate units *a, b, c, ab, ac, bc*, and *abc*. Unlike the RW model, however, the activity level of units is context dependently scaled such that total activity in the system is always equal to the sum of the intensities

of the individual cues present. For example, if we assume that the intensities of all stimuli are individually equal to 1, then when the compound ABC is presented, the combined activity in units *a, b, c, ab, ac, bc*, and *abc* will equal 3. In this manner, each stimulus contributes the same absolute amount of activity to the system whether it is presented alone or in compound.

It is not just the intensity of a stimulus that influences the activation level of a unit, but also the context in which that stimulus is presented. Stimuli will interact with each other perceptually, and the extent of this interaction is reflected by the parameter *r*. For simplicity, discussion here will be restricted to an ideal system in which the intensity of all stimuli is the same and equal to 1, and all stimuli within a system interact with all others to the same degree, *r*. When a single stimulus, A, is presented it will activate a single unit, *a*. In this case, the activation of unit *a*, γ_a, will be equal to the intensity of the stimulus, I_A, which is 1. When A is presented in compound with a second stimulus, B, three units will be activated: *a, b*, and *ab*. Now, the activation of unit *a* will be reduced by the extent to which the two stimuli interact, *r*. Hence, $\gamma_a = (1 - r)$. Stimulus A also contributes to the activation of the *ab* configurational unit and its contribution to γ_{ab} is equal to the reduction in activation of unit *a*, *r*. The interaction between stimuli A and B is reciprocal, and so the activation of unit *b*, γ_b, is also given by $(1 - r)$. Similarly, B contributes *r* to the activation of unit *ab*, so that $\gamma_{ab} = 2r$. In effect, units *a* and *b* lend some of their activation strength to the configurational unit *ab*. It should be apparent that the figures given here for the activation level of each unit correspond to the proportion of elements within the populations A_i, B_i, and the combination of A_b and B_a, in Wagner's REM model. The critical difference is that there are no units that specifically represent the absence of other stimuli when a stimulus is presented alone (i.e. $A_{\sim b}$ or $B_{\sim a}$). Instead, the units that represent the individual stimuli, *a* and *b*, will be more active when the stimuli are presented alone than when they are presented in compound.

The situation is slightly more complicated when three stimuli are presented in compound, but follows the same principle of statistically independent interaction as REM. That is, in compound ABC,

γ_a is reduced by interaction from both B and C and $\gamma_a = (1-r)(1-r) = (1-r)^2$. Stimulus A still contributes r to the activation of unit *ab* (as does B), but activity in *ab* is itself reduced by the interaction with stimulus C. Now, $\gamma_{ab} = r(1-r) + r(1-r) = 2r(1-r)$. Finally, activation will propagate to a unit that represents the configuration of all three stimuli, *abc*. Activation of this unit is $\gamma_{abc} = 3r^2$.

The activity in any unit is given by Equation 9 where k is the number of stimuli contributing to the activation of that unit, and n is the total number of stimuli present. A more general version of Equation 9 is given in Appendix B, which allows γ to be calculated when the intensities of stimuli are not all equal and when stimuli interact with each other to different extents.

$$\gamma = k(1-r)^{n-k} r^{k-1} \tag{9}$$

The number of units that receive activation from k stimuli is given by the binomial coefficient shown in equation 10. For example, in a system with three stimuli (A, B, and C), three units will each receive activation from a single stimulus (*a, b*, and *c*), and another three units will each receive activation from two stimuli (*ab, ac*, and *bc*). In the former case $n = 3$ and $k = 1$, so Equation 10 gives us 3!/1!(3-1)! = 3!/2! = 3. In the latter case, $n = 3$ and $k = 1$, and Equation 10 gives 3! = 2!(3-1)! = 3!/2! = 3. A single unit (*abc*) will receive activation from all three stimuli (3!/3!(3-3)! = 1). It follows from Equations 9 and 10 that the total activity across all units can be calculated by Equation 11, and equals n.

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{10}$$

$$\gamma_{total} = \sum_{k=1}^{n} \binom{n}{k} k(1-r)^{n-k} r^{k-1} \tag{11}$$

Learning within C-AEM proceeds in a manner similar to both RW and REM, but learning is scaled by the activation of each unit. When a stimulus, or stimulus compound, is presented, the

change in the associative strength, *V*, of a particular representational unit is given by Equation 12 where α and β are learning rate parameters associated with the unit and the US, respectively and λ is the magnitude of the US. V_{net} is the expected outcome and is determined by the sum of the products of the activation of each unit and the strength of its association with the US as shown in Equation 13. The activation of a unit affects both its contribution to prediction of the US (V_{net}), and how much is learnt about that unit following a conditioning trial.

$$\Delta V = \alpha\beta\gamma\left(\lambda - V_{net}\right) \tag{12}$$

$$V_{net} = \sum \gamma V \tag{13}$$

C-AEM is computationally a much simpler model than REM whilst retaining the key principle of stimulus interaction. As the number of stimuli that may be presented in compound increases, there is an exponential growth in the number of discrete populations of elements within REM. This growth follows the function $n3^{n-1}$. The growth of representational units within C-AEM is also exponential, but at a much slower rate. The expansion in C-AEM may be derived from binomial theorem (see Appendix C). In a system comprising *n* stimuli, $(2^n - 1)$ units may be activated. Hence, where REM requires 1, 6, 27, 108, and 405 populations of elements to represent systems consisting of 1, 2, 3, 4 and 5 stimuli, respectively, C-AEM requires just 1, 3, 7, 15, and 31 units.

Due to the similarity of C-AEM to REM and the RW model, the two models make similar predictions in a variety of situations. The differences between the models do, however, result in deviations between their predictions in some situations where REM is unable to account for all of the experimental data. To test the predictions of C-AEM and compare them against those made by REM, the RW model, and Configural Theory, a series of computer simulations was conducted.

Application of the Models to Empirical Data

Simulations are presented here of Configural Theory, REM and C-AEM for a selection of situations in which the RW model and Configural Theory make different predictions, and where there is empirical support for the predictions of each. This is not intended to be a comprehensive review of the capabilities of any of the models. Indeed, all of the models considered here make use of a summed error term in their learning rules, which makes it difficult for them to account for changes in the associative strength of stimuli that differ in their associative history when they are conditioned in compound (Rescorla, 2000). Rather, the discussion here is limited to some effects of similarity and generalization which inspired the development of REM, and a closely related patterning discrimination task.

Unless otherwise stated, parameter values were as follows. α was set at .05 for all stimuli and configurational cues for the RW model, all configural units for Configural Theory, all populations of elements for REM, and all representational units for C-AEM. For simulations of all models, λ was equal to 1 and β was .05 when the US was present, and λ was zero and β was .025 when the US was absent[2]. For REM and C-AEM, all *r* values were equal and for C-AEM the intensity parameter, *I*, was set to 1 meaning that Equation 9 could be used to calculate unit activations. For simulations of Configural Theory, $d = 2$. Where a salient contextual cue was included in the simulation, it was treated in the same manner as a stimulus, with $\alpha = .05$ and $I = 1$.

Overshadowing and external inhibition

Pavlov (1927) described an experiment in which two stimuli were paired, in compound, with a US. When each stimulus was presented by itself following this training, a weaker CR was provoked than when they were presented together. This *overshadowing* effect is readily predicted by the RW model and the prediction is easily derived from Equation 2. Compound training will have the effect of increasing Σα and thus reducing the asymptotic value of V_A to something lower than λ. Furthermore, a more salient stimulus will overshadow conditioning to a less salient stimulus to a greater extent than

the more salient stimulus will be overshadowed by the less salient one (see Miles & Jenkins, 1973; Kamin, 1969). Again, this effect is predicted by Equation 2: since the learning rate parameter α reflects the salience of a stimulus, the addition of a highly salient stimulus will have a greater impact on the value of (α_A / Σα) than will the addition of a less salient stimulus. The RW model, however, does not predict *external inhibition*. Pavlov (1927) also observed that if some additional stimulation such as a change in the illumination of the experimental room, or a loud noise from outside, coincided with the presentation of an established CS, the magnitude of the CR was diminished. Similarly, when a stimulus is presented in compound for the first time in blocking experiments, the CR is sometimes smaller than on the preceding conditioning trial when it was presented alone (e.g., Kamin, 1969). Presenting an additional (neutral) stimulus in combination with an established CS should not affect responding to that CS according to the RW model since the associative strength of the compound is simply the sum of the associative strengths of its components.

Configural theory provides a ready explanation for generalization decrements. If an animal has received conditioning with the compound AB, then presentation of stimulus A alone will activate the AB configural unit only to the extent that A is similar to AB. Since the similarity of these two patterns is less than 1 (according to Equation 5, $S_{AB,A}$ = .5), the unit will receive less activation in response to the presentation of A alone than to the presentation of compound AB. A symmetrical effect is predicted when compound AB is presented following conditioning to stimulus A; activation of the A configural unit by compound AB is similarly determined by the similarity of AB to A. Varying the discriminability parameter, *d*, will affect the similarity of patterns, but for all values Configural Theory predicts symmetrical effects of overshadowing and external inhibition.

Brandon, Vogel and Wagner (2000) observed neither the patterns of results predicted by the RW model or by Configural Theory. They trained three groups of rabbits using eye-blink conditioning. For the first group, stimulus A was paired with a paraorbital electrical shock. A second group was trained with compound AB, and the third with compound ABC. Following conditioning, animals in all

three groups received test trials with A alone, and with the compounds AB, and ABC. Both overshadowing and external inhibition were observed; either adding or removing features from each training pattern resulted in a reduction in the conditioned response. These effects were not symmetrical; removing a feature from the training pattern had a greater impact on responding than adding a feature.

Brandon et al's (2000) results are consistent with the predictions of REM. When conditioning is conducted with compound AB, four populations of elements will acquire associative strength. These are context-independent A_i and B_i elements and context-dependent A_b and B_a elements. The relative size of these populations is $(1 - r)$, $(1 - r)$, r and r, for A_i, B_i, A_b, and B_a, respectively. When A is presented by itself, it will activate context-independent A_i elements and context-dependent $A_{\sim b}$ elements. Hence, only the A_i elements are activated by both compound AB and stimulus A alone. The proportion of AB's elements that are also activated by A is $\frac{1}{2}(1 - r)$ because none of B's elements are activated by stimulus A. Conversely, whenever a feature is added to a pattern, it will result in the replacement of a fixed proportion, r, of the elements activated by that pattern. The elements of A that are also activated by compound AB are again the context-independent A_i elements, and the portion of A's elements that are activated by compound AB is $(1 - r)$. Generalization between AB and ABC follows similar rules. Generalization of associative strength from AB to ABC will again be equal to $(1 - r)$; addition of a stimulus results in replacement of a fixed proportion of the total elements of the original pattern. Generalization of associative strength from ABC to AB is, however, predicted to be $\frac{2}{3}(1 - r)$ because only two of the three stimuli compound ABC are also present in compound AB. In all cases removal of a feature is expected to have a greater effect than the addition of a feature.

The predictions that C-AEM makes concerning the relative size of the effects of overshadowing and external inhibition are not as straightforward as those of REM. Rather, they depend on the number of features in the training and testing patterns. C-AEM makes the same predictions as REM about external inhibition; adding a stimulus will always reduce activity in units by

the proportion *r*. Removing a stimulus, however, does not simply result in fewer units being activated, but also changes the activation level of those units. The top-left panel of Figure 1 shows how generalization between A and AB varies with *r*. For all values of *r*, the effect of over-shadowing is greater than that of external inhibition. The top-right panel of Figure 1 shows corresponding predictions concerning generalization between A and ABC. Here, we can see that for some values of *r* between about .3 and .5, the difference in the size of the effects of over-shadowing and external inhibition is quite small. The bottom two panels of Figure 1 show predictions for generalization between ABC and either AB (left panel) or ABCD (right panel). In both cases, there are values of *r* for which the effect of external inhibition is predicted to be greater than that of overshadowing.

Figure 1 about here

Asymmetrical generalization has been reported in several experiments with rats (González, Quinn & Fanselow, 2003; Bouton, Doyle-Burr & Vurbic, 2012) and humans (Glautier, 2004; Wheeler, Amundson & Miller, 2006; Thorwart & Lachnit, 2010). Other authors have, however, observed symmetrical effects of overshadowing and external inhibition in very similar situations (Young, 1984 cited in Pearce 1987; Rescorla, 1999; Thorwart & Lachnit, 2009). Perhaps then, it is premature to suggest that the experimental evidence provides particularly strong support for any one of these theories over any other. I am, however, aware of no evidence in support of C-AEM's prediction that, under some conditions, the effect of external inhibition should be greater than that of overshadowing. REM and C-AEM also make the seemingly unreasonable prediction that adding a feature to a pattern will result in the same decrement in generalization regardless of the number of features of which that pattern is composed. Brandon et al's (2000) results, however, support this prediction: two groups of

rabbits given conditioning with stimulus A or compound AB and then tested with compounds AB or ABC, respectively, showed equivalent decrements in responding (16% vs. 18%).

Simple Summation

If two stimuli, A and B, that have been separately paired with a US are then presented together, responding to the compound AB is sometimes observed to be greater than to either of the individual stimuli (e.g., Whitlow & Wagner, 1972). This *summation* effect is predicted by the RW model because first, the associative strength of a stimulus compound is assumed to be equal to the sum of the associative strengths of its constituent stimuli (i.e. $V_{AB} = V_A + V_B$) and second, the relationship between associative strength and response strength is assumed to be at least ordinal.

Configural Theory struggles to explain simple summation. Following A+ B+ training, responding to the compound AB will depend upon generalization of associative strength from the A and B configural units. According to Equation 5, the similarity of AB to each of the individual cues is .5. Hence, half of the associative strength of each stimulus will generalize to AB and the net associative strength of AB will be the average of the associative strengths of A and B. The failure to predict summation might not, however, be catastrophic for configural theory. It should be noted first that summation is not a ubiquitous effect. Although it has been observed in some experiments (e.g., Aydin & Pearce, 1995, 1997; Kehoe, 1986; Kehoe, Horne, Horne & Macrae, 1994), there are others where it has not (e.g., Aydin & Pearce, 1995, 1997; Kehoe, Horne, Horne & Macrae, 1994). There are also non-associative explanations of the summation effect, such as *stimulus intensity dynamism* (e.g., Hull, 1949) and *disinhibition of delay* (e.g., Pavlov, 1927).

Configural theory does not have to rely on non-associative mechanisms to explain summation in all situations. In one experiment, Pearce, George & Aydin (2002) gave rats training in which two stimuli, A and B, were each individually paired with food, as was the compound CD (A+ B+ CD+). At test, responding was greater to compound AB than to CD – summation was observed. In a second experiment, the same comparison was made between-subjects. One group received A+ B+ training,

and a second were trained with just AB+ trials. At test there was no difference in the rate of responding during presentations of compound AB for these two groups. Inclusion of CD+ trials influenced summation of responding to A and B. One explanation for this effect that Pearce et al considered concerned the nature of the stimuli used, and the effect that this might have on generalization between stimuli. In these experiments, A and C were visual stimuli, whereas B and D were auditory stimuli. Pearce et al suggested that stimuli from the same modality might share some common features which are not shared between stimuli belonging to different modalities. Hence, A, B, C, and D may be conceptualized as *ax*, *by*, *cx*, and *dy*, and the compounds AB and CD as *abxy* and *cdxy*. Due to the influence of generalization on the asymptotic associative strengths of the various configural units, and generalization of associative strength from *ax*, *by*, and *cdxy* to abxy at test, configural theory predicts the net associative strength of *abxy* (1.1λ) will be greater than that of *cdxy* (λ) for the within-subject comparison. For reasons explained earlier, no summation would be expected following simple A+ B+ training since half of the associative strength of *ax* and of *by* will generalize to *abxy*.

Nevertheless, several experiments have demonstrated summation following simple A+ B+ training (e.g., Hendry, 1982; Kehoe, 1986; Konorski, 1948, Thein, Westbrook & Harris, 2008). Even in these circumstances, however, it is possible for Configural Theory to explain summation, if it is assumed that the experimental context is of relatively high salience. This seems to be a reasonable assumption in some cases at least, for example where aversive Pavlovian conditioning is conducted over an appetitive instrumental baseline (Hendry, 1982), or conditioning is conducted in restrained rabbits (Kehoe, 1986) or restrained dogs (Konorski, 1948). Where the context is salient, simple A+ B+ training may be re-described as AX+ BX+ Xø where X is the salient context, and the test compound is ABX. If the saliences of A, B, and X are equal, then at the asymptote of conditioning the net associative strengths of patterns AX and BX will be equal to λ and that of X will be 0. The associative strengths of the configural units, however, will be as follows: AX = BX = 1.33λ; X = –1.33λ. Since the similarity of ABX to AX and BX is high (.66), and relatively little inhibitory associative strength will generalize to ABX from X due to their low similarity (.33), the net associative strength of the compound is predicted to

be 1.33λ. Pearce, George, Redhead, Aydin and Wynne (1999; see also Aydin & Pearce, 1997; Pearce, Redhead & George, 2002) reported a related effect in pigeon autoshaping. They manipulated generalization between stimuli A and B and the test compound AB by changing the salience of the background illumination of the television screen on which they were presented. Summation was observed when the background was white and was also illuminated throughout the inter-trial-interval, but not when it was black. The results of simulations of Configural Theory are shown in the top panel of Figure 2. Stimuli A and B were individually paired with a US until each achieved asymptotic associative strength, and test trials were then given with the individual stimuli and their compound (A+ B+; A? B? AB?). Summation was predicted only when a salient contextual cue was added (AX+ BX+ Xø; AX? BX? ABX?).

Figure 2 about here

Configural Theory can also predict simple summation if generalization between patterns is increased by giving the *d* parameter a value lower than 2. For example, when *d* = 1.5, the net associative strength of compound AB following asymptotic conditioning with A and B will be 1.19λ. Reducing the value of *d* does, however, cause some problems for Configural Theory. For example, when *d* < 2 Configural Theory predicts that complex negative patterning discriminations of the form A+ B+ C+ AB+ AC+ BC+ ABCø are insoluble, which is not the case (Redhead and Pearce, 1995). Increasing *d* above 2 decreases generalization and results in compound AB having lower net associative strength than A or B alone.

REM and C-AEM make the same predictions as each other concerning simple summation following conditioning with patterns that share no common features. In this situation, generalization from the training patterns to the test pattern is determined purely by the *r* parameter. For REM, during

conditioning with stimulus A, content independent elements A_i and context-dependently inhibited elements $A_{\sim b}$ will accrue associative strength. During presentations of test compound AB, these latter elements will not be active, and generalization of associative strength from A to AB will be based upon the proportion of context independent elements, $s = (1 - r)$. Associative strength will generalize from stimulus B to compound AB in the same way, and the compound will have a net associative strength of $2(1 - r)$. In C-AEM, stimulus A will only activate its own representational unit, *a*, and that unit will gain associative strength until $V_A = \lambda$. The activation of unit *a* (and unit *b*) when compound AB is presented may be calculated using Equation 9 and will again be equal to $(1 - r)$. The lower panel of Figure 2 shows the results of the predictions of REM and C-AEM for a summation experiment where stimuli A and B were presented alone and in compound following conditioning with A and B. The models predict that summation will occur when interaction between the stimuli is low ($r < .5$). For high values of the *r* parameter ($r > .5$), they predict less responding to compound AB than to either A or B individually. Hence, all three models can accommodate the results of experiments which have demonstrated response summation using stimuli drawn from different modalities (e.g., Whitlow & Wagner, 1972) and those which have failed to find evidence of summation within a single stimulus modality (e.g., Rescorla & Coldwell, 1995), if we assume that $r < .5$ in the former case, and $r \approx .5$ in the latter case. REM and C-AEM also predict lower responding to a compound when the replacement parameter is very high, an effect that has been observed by Aydin & Pearce (1995, 1997).

Differential Summation

Pearce, Aydin & Redhead (1997) gave pigeons autoshaping training in which presentations of three visual stimuli, A, B, and C, were paired with food. For one group of pigeons, these stimuli were presented individually (A+ B+ C+), whereas for a second group, they were presented in pairs (AB+ AC+ BC+). Following this training, responding to the compound ABC was assessed in each group. Responding during these test trials was slower in the group trained with the individual stimuli. This result is predicted by Configural Theory because the similarity of ABC to the compounds AB, AC, and BC is greater than the similarity of ABC to the individual stimuli A, B, and C. Although generalization

between the pairs of stimuli will result in the configural units for AB, AC, and BC each having an asymptotic associative strength of .66λ, two-thirds of the associative strength of each compound will generalize to the compound ABC, resulting in a net associative strength of 1.33λ. When the three stimuli are trained individually the asymptotic associative strengths of the configural units A, B, and C, will be λ. Because only one third of the associative strength of each will generalize to ABC the net associative strength of the test pattern is predicted to be λ. The top-left panel of Figure 3 shows the predicted net associative strength of compound ABC for the two groups in Pearce et al's experiment (A+ B+ C+; ABC? or AB+ AC+ BC+; ABC?).

The RW model makes the opposite prediction. It supposes that training with the individual stimuli will result in each gaining an asymptotic associative strength of λ. When the stimuli are trained in compound, however, Equation 2 predicts that the asymptotic associative strength of each stimulus will be .5λ. Because the net associative strength of the compound is the simple arithmetic sum of the associative strengths of A, B, and C, RW predicts that ABC will have an associative strength of 3λ following individual training, but of only 1.5λ following compound training. Myers et al (2001) replicated Pearce et al's (1997) experiment in rabbits using eyeblink conditioning with three stimuli drawn from different modalities (auditory, visual, and tactile), and found greater levels of responding to ABC in the group that received A+ B+ C+ training than in the group the received AB+ AC+ BC+ training, consistent with the predictions of the RW model.

Configural Theory can accommodate Myers et al's results if it is assumed that a salient contextual cue is present (AX+ BX+ CX+ Xø; ABCX? or ABX+ ACX+ BCX+ Xø; ABCX?). Results of simulations of Configural Theory with the inclusion of a contextual cue are shown in the top-right panel of Figure 3. During AX+ BX+ CX+ Xø training, there will be some generalization between the compounds ($S_{AX,BX}$ = .25), but considerably more generalization between each compound and the context ($S_{AX,X}$ = .5). This means that the configural units for AX, BX, and CX will have asymptotic associative strengths of 1.33λ whereas the unit for X will have an associative strength of -2λ. Half of

the associative strength of AX, BX, and CX will generalize to ABCX, but only one quarter of that of X will. Hence, the net associative strength of the test compound is 1.5λ. For ABX+ ACX+ BCX+ Xø training, there will be considerably more generalization between the compounds ($S_{ABX,ACX}$ = .44) and less generalization between the compounds and the context ($S_{ABX,X}$ = .33). At test, three quarters of the associative strength of each compound and one quarter of the associative strength of the context will generalize to ABCX, meaning that its net associative strength will be 1.29λ. Configural units ABX, ACX, and BCX will end up with associative strengths of .64λ and X will have an associative strength of -.64λ. Configural Theory also predicts that compound ABC will have greater net associative strength following A+ B+ C+ training than following AB+ AC+ BC+ training if the *d* parameter is reduced significantly below 2, but it is difficult to justify increasing generalization when it prevents the model from solving some discrimination problems. Increasing the value of *d* above 2 does not affect the ordinal predictions of Configural Theory.

Figure 3 about here

When training patterns share common features, the predictions of REM and C-AEM diverge. Predictions from REM may be derived quite straightforwardly. Since replacement of A elements by different stimuli (i.e. B and C) is statistically independent, there are nine distinct populations of A elements to be considered; each population is either context independent, context-dependently inhibited, or context-dependently activated with respect to B and with respect to C. These nine populations and their relative sizes are enumerated in Table 1.

Table 1 about here

For animals trained with the individual stimuli A, B, and C, none of the added elements (i.e., A_b, A_c,, $A_{b\sim c}$, $A_{c\sim b}$, A_{bc}) will be activated during training, and generalization from A to the test compound ABC will rely solely on the elements that are context independent with respect to both B and C (A_i). Thus, generalization from A to ABC will equal $(1 - r)^2$. Associative strength will also generalize from B and C to ABC in the same proportions. When $r = 0$, $V_{ABC} = 3\lambda$. As the amount of replacement increases, generalization will decrease exponentially until the net associative strength of ABC will equal zero when $r = 1$. For animals trained with compounds AB, AC, and BC, generalization from stimulus A to ABC will rely not only on the context independent A_i elements, but also on the context dependent elements that are commonly activated by compounds ABC and either AB or AC (i.e. A_b and A_c). The combined size of these latter populations of elements is given by $2r(1 - r)$. The relationship between the value of this term and *r* is not monotonic. Instead, it increases with *r* in the range $0 \leq r \leq .5$ until it reaches a maximum value of .5 but decreases as *r* increases in the range $.5 \leq r \leq 1$. This means that as *r* increases in value, a decline in the size of the population of context-independent A_i elements will, to some extent, be offset by an increase in the number the context-dependent A_b and A_c elements. The interplay between these different populations of elements means that for some intermediate values of *r*, REM predicts that the test compound ABC will have greater net associative strength following AB+ AC+ BC+ training than following A+ B+ C+ training (see the middle panel of Figure 3).

In C-AEM, the activation level of a unit affects both its contribution to the net associative strength of a pattern, and also the change in the associative strength of that unit following a conditioning trial. Because of this, the contribution of units to the net associative strength of a compound in C-AEM can differ from the contribution of corresponding populations of elements in REM. Following conditioning with the individual stimuli (A+ B+ C+), only the representational units for the stimuli themselves (*a, b*, and *c*) will accrue associative strength in C-AEM. For all values of *r*, the associative strength of each unit will approach λ at the asymptote of conditioning. When compound

ABC is presented at test, each of these units will have an activation of $(1 - r)^2$, and hence the net associative strength of the compound will be $3(1 - r)^2$. This is the same prediction as REM. During compound conditioning, however, units *ab, ac*, and *bc* will also acquire associative strength. Because the activation level of each unit acts as a rate parameter in Equation 12, the distribution of associative strength between the representational units activated by a compound is influenced by *r*. Within each compound, the proportion of associative strength that will accrue to the configurational cue will be equal to *r*: $V_{ab} / (V_a + V_b + V_{ab}) = r$. At test, the activation of the configurational cue *ab* will be $2r(1 - r)$, and the combination of these two influences of *r* means that the unit makes very little contribution to the net associative strength of compound ABC for low values of *r*. The results of simulations of C-AEM are shown in the bottom panel of Figure 3. Despite some differences in the predictions of REM and C-AEM, they agree that the net association strength of ABC following compound training will be lower than following individual training for values of *r* less than about .35. The opposite will be true for values of *r* greater than about .4.

Differential External Inhibition

Pearce, Adam, Wilson and Darby (1992) also compared responding to the compound ABC in two groups of pigeons given slightly different training involving the three stimuli. The first group were trained on a discrimination involving the individual stimuli in which A and C were each individually paired with food whereas B was not (A+ Bø C+). The second group got the same training except that B was present on all trials (AB+ Bø BC+). Responding to ABC at test was greater in the latter group, a result consistent with Configural Theory because ABC is more similar to AC and BC than it is to A or C. In contrast, the RW model predicts that there should be no difference in responding to ABC by the two groups because A and C should each reach an asymptotic associative strength of λ, while B should have an associative strength of zero. Kundey, Tronson and Wagner (2002; unpublished data cited in Wagner, 2003) found the same pattern of results as Pearce et al in a rabbit eye-blink experiment using stimuli from different modalities.

In order that Configural Theory may explain effects such as summation and differential summation, we may assume that aversive conditioning procedures result in the presence of a salient contextual cue throughout the experimental session. Given Kundey et al's (2002) results, it is significant, therefore, that the inclusion of a salient contextual cue does not affect Configural Theory's ordinal predictions in the case of this experimental design. Configural Theory predicts that the net associative strength of ABC will be .67λ following A+ Bø C+ training and 1.33λ following AB+ Bø BC+ training. Corresponding values when there is a salient contextual cue are λ and 1.5λ, respectively. Wagner (2003) showed that REM also makes the same (correct) prediction for all values of *r* except for zero and 1. In the former case, REM is equivalent to the RW model. It is encouraging that C-AEM makes the same predictions as REM for this design where experimental findings have been consistent, just as it does for others where empirical results differ between laboratories. Both models predict that the net associative strength of ABC will be $2(1 - r)^2$ following A+ Bø C+ training and $2(1 - r)$ following AB+ Bø BC+ training. Because $(1 - r) \leq 1$, the former value is smaller than the latter except when $r = 0$ or $r = 1$, when they are equal.

Complex Negative Patterning

For a feature negative discrimination of the form A+ ABø, the addition of a common element, C, to each pattern (AC+ ABCø) retards acquisition of the discrimination (Pearce & Redhead, 1993; Rescorla, 1972). Configural Theory predicts this effect because the similarity of (and therefore, generalization between) the reinforced and non-reinforced patterns is increased by the addition of a common element. Redhead and Pearce (1995) further examined the effects of similarity on discrimination learning using a complex negative patterning design in which three stimuli, A, B, and C, were each paired with food when presented individually or in pairs, but were paired with no food when all three were presented together (A+ B+ C+ AB+ AC+ BC+ ABCø). Configural Theory predicts that acquisition of the discrimination between the individual stimuli (A, B, C) and the triplet (ABC) should proceed more rapidly than acquisition of the discrimination between the pairs of stimuli (AB, AC, BC) and the triplet. The top-left panel of Figure 4 shows the results of a simulation of Configural

Theory trained on this complex negative patterning discrimination. These predictions were confirmed by Redhead and Pearce in a series of autoshaping experiments with pigeons in which the three stimuli were all drawn from the visual modality.

Figure 4 about here

According to the RW model, compounds AB, AC, and BC will provoke larger CRs than the individual stimuli A, B, and C due to summation of associative strength. This model, therefore, predicts the opposite pattern of results to that observed by Pearce and Redhead (1993) and by Rescorla (1972). In order to solve the discrimination and reduce responding to the non-reinforced compound ABC, the model must assume that at least some stimulus conjunctions generate a unique configurational cue which will gain inhibitory associative strength. Slightly different predictions may be derived dependent upon whether configurational cues are generated by each conjunction, or by ABC alone, but these differences do not affect the ordinal prediction that the net associative strength of AB, AC, and BC will be higher than that of A, B, and C alone (before learning reaches asymptote). Myers et al (2001) found support for the predictions of the RW model, and not for Configural Theory, when they trained rabbits using eyeblink conditioning and stimuli from three different modalities. The addition of a salient contextual cue does not substantially affect the predictions of Configural Theory, as can be seen in centre-top panel of Figure 4. Responding to the compounds AB, AC, and BC is predicted to be greater than to the individual stimuli very briefly early in training, but Configural Theory does not provide a convincing match to Myers et al's results. Similarly, changing the value of the *d* parameter does not allow Configural Theory to explain Myers et al's results. When $d < 2$, the model predicts that the discrimination is insoluble. For all values of *d* greater than 2, the modified version of Configural Theory predicts greater responding to reinforced patterns than to the non-reinforced compound ABC. As *d*

increases, the difference in responding to A, B, and C and the compounds AB, AC, and BC is predicted to decrease, so that for large values (around 10), responding to all reinforced patterns should be approximately equal. This pattern of results was observed by Kinder and Lachnit (2003) using a human eye-blink conditioning procedure. For no value, however, is responding to A, B, and C predicted to be lower than to AB, AC, and BC; the model is incapable of predicting Myers et al's results.

Simulations of REM reveal that it cannot generate predictions consistent with Redhead and Pearce's (1995) results. The middle row of Figure 4 shows simulation of REM for *r* values of .2, .4., and .8. Additional simulations, not shown here, confirmed that for all values of *r* greater than zero, REM can solve the discrimination, but at no point in any simulation was the net associative strength of compounds AB, AC, and BC lower than that of the individual stimuli A, B, and C. Further simulations found that addition of a salient contextual cue did not affect the ordinal predictions of REM (or C-AEM) for the complex negative patterning discrimination task, or any of the other effects considered in this paper.

C-AEM's ordinal predictions are affected by the value of the *r* parameter. For very high values of *r*, its predictions are similar to those of REM, but for intermediate values it much more closely resembles Configural Theory. It can predict the results of the pigeon experiments if we assume that *r* equals about .4 or .5 for those experiments; this seems a reasonable assumption since C-AEM (and REM) predict the correct pattern of responding in summation and differential summation experiments for these values. However, when *r* is much lower (around .2, the value which Wagner, 2003, suggests might be appropriate for modelling the results of rabbit eye-blink experiments) the pattern is less clear. For a quite extended period early in training, C-AEM predicts Myers et al's (2001) pattern of result, but later flips over to match those of Redhead and Pearce (1995). Unfortunately, we do not have any data on the effects of extended training on the complex negative patterning task in rabbits, but it could be that Myers' data reflects an early training period; their rabbits showed only a moderate

discrimination between the patterns at the end of training with a high level of responding to the non-reinforced compound ABC.

A study of causal learning in humans gives some support to this interpretation of Myers et al's (2001) results. In two experiments, Redhead (2007) trained participants on the same complex negative patterning discrimination. In his first experiment, all stimuli were coloured dots, but in the second, they came from three different modalities. At the end of training, Redhead reported the same pattern of casual ratings in both cases: ratings for A, B, and C were greater than those for AB, AC, and BC, consistent with Configural Theory. Examination of the result from the participants trained with multimodal stimuli suggests, however, that for a couple of blocks of trials towards the beginning of training rating were higher for AB, AC, and BC than for A, B, and C. Although this numerical difference did not reach conventional levels of statistical significance, it persisted beyond the point where ratings for ABC were significantly lower than for the reinforced patterns. In a third experiment, Redhead also found simple summation of responding for multimodal stimulus compounds, but not unimodal ones. Hence, for multimodal stimuli the results of the summation test and the course of acquisition of the complex negative patterning discrimination matched the predictions of C-AEM for a value of *r* around .2, whereas for unimodal stimuli the results were consistent with the predictions of C-AEM for values of *r* in the region of .4 – .5.

Parity

Pearce et al (2008) trained pigeons on a slightly different patterning discrimination of the form A+ B+ C+ ABø ACø BCø ABC+ in which the individual stimuli and the compound of all three were paired with food whereas the compounds comprising pairs of stimuli were not. I have previously referred to this as a parity discrimination (George, 2018) because patterns containing odd and even numbers of features are treated differently from each other. Pearce et al found that when stimuli A, B, and C were very similar to each other – three white circles differentiated only be their position on a television monitor – pigeons were unable to solve this discrimination problem. They responded most rapidly to

A, B, and C, and least rapidly to the reinforced compound ABC; responding to the non-reinforced compounds AB, AC, and BC was at an intermediate rate. These results matched the predictions of Configural Theory, which are shown in the top-left panel of Figure 5. A different pattern of results was found when the stimuli were more different to each other. When A, B, and C were collections of dots of different colours (red, green, and blue, respectively), the pigeons learned to peck at the reinforced patterns A, B, C, and ABC, and withhold responding to AB, AC, and BC. However, they responded more rapidly to stimuli A, B, and C than to compound ABC. It appears that reducing the similarity of the individual stimuli had the effect of reducing generalization between the patterns composed of them, which facilitated the acquisition of the discrimination. Increasing the value of *d* in Equation 7 has the same effect and allows Configural Theory to predict the results of this experiment, as shown in the top-centre panel of Figure 5. There is no reason to suggest that there was a salient contextual cue present in either of Pearce et al's experiments and addition of one does not allow Configural Theory to solve the discrimination when $d = 2$.

Figure 5 about here

REM is capable of solving the parity discrimination, but it incorrectly predicts that responding should be greater for compound ABC than for A, B, or C alone (see the middle row of Figure 5). This is true for all values of *r* greater than zero. When $r = 0$ (not shown), REM fails to solve the discrimination but predicts that the order of rates of responding is as follows: ABC > AB/AC/BC > A/B/C. Hence, REM is unable to predict the results of either of Pearce et al's (2008) experiments for any value of *r*.

C-AEM predicts the same pattern of results as REM for very high values of *r* (bottom-right panel of Figure 5). For intermediate values, around .4 (bottom-centre panel of Figure 5), it correctly predicts the results of Pearce et al's (2008) experiment using coloured dots. There is a brief period

early in training where C-AEM predicts summation (ABC > AB/AC/BC > A/B/C), but this is also true of Configural Theory and for both models responding to A, B, and C is predicted to be greater than to ABC soon after responding to the non-reinforced patterns begins to decline. It is possible that stimuli which are identical except for their location may interact with each other perceptually rather less than different stimuli within the same modality. When *r* is small (.2), C-AEM still solves the discrimination, but extremely slowly (bottom-left panel of Figure 7; note that the scale on the x-axis is an order of magnitude different than for the other simulations of C-AEM in the same figure). Here, there is an extended period during which the model predicts a similar pattern of responding to that observed by Pearce et al in the experiment using white circles as stimuli (A/B/C > AB/AC/BC > ABC), although the model's predictions do not match the performance of the pigeons as closely as do the predictions of Configural Theory.

To the best of my knowledge, this type of parity discrimination has been employed only in the experiments reported by Pearce et al (2008) and in one other study (George, 2018). I trained human participants on simultaneous negative (A+ B+ C+ ABø ACø BCø ABC+) and positive (Dø Eø Fø DE+ DF+ EF+ DEFø) parity discriminations in a predictive learning task. For some participants, each of the stimuli was a circle that consistently appeared in a specific location on a computer monitor (the six stimuli were arranged at the vertices of an imaginary hexagram so that ABC and DEF were the corners of two equilateral triangles), and all of the circles were of the same colour. For other participants the six stimuli were quite different from each other, based on those used by Pearce and George (2002) in a pigeon auto-shaping experiment. On each trial, participants were asked to rate the likelihood that the visual pattern would be followed by the presentation of a tone (+) on a scale that ranged from 1 = *very unlikely* to 9 = *very likely*. Both groups of participants solved the discrimination problems, but stimulus similarity affected predictive ratings for the individual stimuli and the compounds of three stimuli (ABC/DEF). When stimulus similarity was low, participants made more extreme ratings to the individual stimuli than to the three-stimulus patterns (i.e. A/B/C > ABC; D/E/F < DEF). The group trained with very similar stimuli showed the opposite pattern of results (i.e. A/B/C < ABC; D/E/F > DEF).

The results of simulations of Configural Theory, REM, and C-AEM are shown in Figure 6 for the human parity discrimination task. For simplicity, only the data from the negative parity discrimination are shown; data from the positive and negative parity discriminations are rotations of each other around the zero point of the y-axis. For these simulations, β was .05 both when the US was present and when it was absent, and λ was -1 when the US was absent. These values were chosen because in human causal learning experiments the presence and absence of the outcome are usually events of equal salience and likelihood (Livesey, Thorwart & Harris, 2011). The predictions of Configural Theory are consistent with the results from the group of participants trained with low similarity stimuli when $d > 2$ (top-centre panel). When $d = 2$, Configural Theory cannot solve the discrimination (top-left panel). For all values of r, REM predicts that more extreme ratings should be given to the three-element patterns that to the individual stimuli (middle row), the effect shown by participants trained with high-similarity stimuli. Only C-AEM can predict both patterns of results. The low-similarity stimuli differed across three perceptual dimensions (colour, shape, and orientation), although they were all visual. Pearce and George (2002) found that pigeons trained on a complex negative patterning discrimination with these stimuli behaved in a similar manner as Redhead and Pearce's (1995) pigeons trained with stimuli that differed along a single dimension (colour). Hence, it is reasonable to set the r parameter to a value in the range .4 – .5 for which C-AEM correctly predicts the results of other pigeon auto-shaping experiments. As can be seen in the bottom-centre panel of Figure 6, the predictions of C-AEM with $r = .4$ match the pattern of rating for the low-similarity stimuli. For high values of r (.8), C-AEM predicts the pattern of results obtained from participants trained with high-similarity stimuli (bottom-right panel). It may be fair to suppose that perceptual interaction was abnormally high for the high-similarity stimuli. Because the individual elements of the patterns were identical, defined only by their location, their geometric arrangement may have been very salient. A, B, and C alone were simply points in space; AB, AC, and BC were pairs of points lying along lines a 120°, 60°, or 0° from horizontal, respectively; ABC was three points at the corners of an implied equilateral triangle. When the individual stimuli were less similar their identities may have been rather more

salient than their spatial arrangement, resulting in less perceptual interaction (but still more than for multi-modal stimuli). Simulations of C-AEM reproduced another aspect of the experimental data; C-AEM predicts faster learning when $r = .8$ than when $r = .4$. Although there was no statistically significant difference in the overall rates of learning of the two groups, there was a trend for the group trained with high-similarity stimuli to learn more rapidly than those trained with low-similarity stimuli.

In order for C-AEM to explain both Pearce et al's (2008) and my (George, 2018) results, it is necessary to assume that stimuli that differ only in their location interact with each other to a much lesser extent for pigeons that for people. Points on a computer monitor might be seen as the vertices of shapes by people, but not by pigeons. There is some evidence that this is true, or at least that vertices and edges are processed differently by the two species. Biederman (1987), for example, found that people's recognition of simple line drawings of objects was impaired to a much greater extent by the deletion of vertices than by the deletion of edges. The opposite effect was found in pigeons by Rilling, De Marse and La Claire (1993; see Qadri & Cook, 2015, for a discussion of divergences between avian and mammalian visual cognition).

Summary

The context-dependent added-elements model presented here is computationally less complex than Wagner's (2003) replaced elements model while retaining many of the properties of REM. C-AEM is able to accommodate conflicting findings concerning summation and differential summation from experiments involving pigeon autoshaping (e.g., Pearce et al, 1997) and rabbit eyeblink conditioning (e.g., Myers et al, 2001), and also the consistent effect of differential external inhibition in these two preparations (Pearce et al, 1992; Kundy et al, 2002) in a similar manner to REM. In addition to its relative simplicity, C-AEM has the strength that it provides a better overall fit to data from experiments involving complex patterning discriminations than does either REM or Pearce's Configural Theory. For both complex negative patterning and parity discriminations, REM predicts a consistent (ordinal) pattern of results for all values of r greater than zero. Configural Theory provides

a very close match to data from Pearce et al's (2008) parity experiments, but cannot accommodate the differential results of complex negative patterning experiments (Redhead & Pearce, 1995; Myers et al, 2001) or the effects of similarity on human parity discrimination learning (George, 2018).

Other models of learning

The primary purpose of this paper was to explore whether the configurational cue, or added elements, approach of Spence (1952) and Wagner and Rescorla (1972) was able to account for the variety of data concerning the effects of similarity on generalization and discrimination learning as well as Wagner's (2003) replaced elements model. The clear answer is yes, and better, if the activity of representational units is context-dependent. There are, however, other elemental models of learning which may be able to account for some of the effects of stimulus modality on discrimination learning. Harris and Livesey (2010; Thorwart, Livesey & Harris, 2012) described an Attention-Modulated Associative Network (AMAN) model in which the activation of representational elements undergoes a normalization process which is affected by stimulus similarity. They have suggested that stimuli from the same modality are more similar to each other than are those from different modalities; similarity produces the modality effect. AMAN is able to reproduce the effect of stimulus modality in some circumstances. Redhead and Pearce (1993) trained pigeons on two concurrent patterning discriminations where all of the stimuli were coloured dots. The first discrimination was of the standard form: A+ B+ ABø. In the second, a common feature was present on all trials: CD+ CE+ CDEø. Addition of the common element increased similarity between reinforced and non-reinforced patterns and retarded acquisition of the discrimination, as predicted by Configural Theory. Bahçekapılı (1997; described in Myers et al, 2001) found the opposite pattern of results when he trained two groups of rabbits on feature negative discriminations with (AC+ ABCø) or without (A+ ABø) a common feature using multi-modal stimuli. Redhead and Curtis (2013) replicated this moderating effect of stimulus modality on the ability of a common feature to retard or enhance discrimination learning in a human contingency learning experiment. They also conducted simulations of AMAN, which showed

that by manipulating the similarity of the stimuli, the model could predict their results. However, in a direct test of the AMAN's predictions, I manipulated the similarity of stimuli across four patterning tasks and in no case could AMAN accommodate the results (George, 2018).

McLaren and Mackintosh (2000, 2002; McLaren, Kaye & Macktinosh, 1989) presented a real-time elemental model of learning. The predictions that this model makes concerning the addition of a common element to a patterning or feature-negative discrimination depends upon how much learning takes place on each trial. When little is learned, the model's predictions match those of the RW model, but when the amount learned is high, they match the predictions of Configural Theory. Conflicting results of experiments using multimodal and unimodal stimuli (or rabbit eyeblink conditioning vs. pigeon autoshaping) may, therefore, be the result of differences in learning rates. It is, however, difficult to compare learning rates between experiments and few experiments have made a direct comparison. Redhead and Curtis (2013) found no difference in the rate of learning between participants trained with multimodal or unimodal stimuli even though the relative rates at which they learned a simple negative patterning discrimination (A+ B+ ABø) and one with the addition of a common feature (CE+ DE+ CDEø) was affected by stimulus modality. I also found no difference in the overall rate of learning between groups of participants that showed the patterns of results predicted by the RW model and by Configural theory for a variety of patterning discrimination tasks (George, 2018). In fact, across four experiments there were numerical differences in learning rate in the direction opposite to that predicted by McLaren and Mackintosh. Recently, Kokkola, Mondragón & Alonso (2019) described a novel elemental connectionist network model, which can account for a wide range of learning phenomena. It is possible that this model may also accommodate the results discussed here, but that so far remains to be confirmed.

Concluding comments

REM and C-AEM can, with slightly different levels of success, reproduce the predictions of both the RW model and Configural Theory. Their flexibility suggests that differences in the predictions

of the latter two models are not simply the results of differences between elemental and configural stimulus processing. Indeed, Ghirlanda (2015) has demonstrated that, under conditions met by existing configural models, it is always possible to construct a configural model equivalent to a given elemental model. Thorwart, Uengoer, Livesey and Harris (2017) have argued that the critical differences between the RW model and configural theory concern normalization of activity within the models and context-dependency of stimulus processing. In Configural Theory, activity of units within the input network are normalized so that only one configural unit is ever fully activated. The activation of configural units is also context-dependent; configural unit A is fully activated by stimulus A, but only partially activated by compound AB. In the RW model, conjunctions of stimuli will result in the generation of a unique configurational cue, but this is in addition to the elemental representations of the component stimuli which are otherwise unaffected by the presence or absence of others. Hence, stimulus representations in RW are both context-independent and non-normalized. In both REM and C-AEM, the activation of individual elements, or the activation level of a unit, is dependent upon the context in which a stimulus is presented, but each stimulus always contributes the same amount of activity to the system; these models are context-depended, but non-normalized. The Inhibited-Elements Model (IEM) described by Wagner and Brandon (2001) as an elemental equivalent to Configural Theory is a normalized, context-dependent elemental model. By varying the degree of normalization with the IEM and comparing its predictions with those of REM and Configural Theory, Thorwart et al (see also Thorwat & Lachnit, 2020) were able to independently evaluate the importance of normalization and context-dependency to the models' ability to predict acquisition of positive and negative patterning discriminations. They concluded that a low-level of context dependency was the critical factor in replicating their results rather than either normalization or a difference in elemental vs. configural processing. Although not presented here, the predictions of C-AEM match those of REM for Thorwart et al's task. This should not be surprising since C-AEM shares those same properties of non-normalization and context-dependence.

Given that REM and C-AEM are both non-normalized and context-dependent, we must consider why they make different predictions in some instances. There are three key differences between the models. First, in C-AEM (and the RW model and Configural Theory), stimuli are represented by units, whereas in REM representations consist of collections of elements or microfeatures. It is unlikely that this difference is responsible for variations in the models' behaviour. Glautier (2007) suggested that an alternative way of simulating REM was to treat each population of elements as a single unit which could be either on or off, and to scale changes in associative strength of these units by a parameter, ω, reflecting the relative size of the populations in REM. This approach yields the same results as simulations in which stimuli are represented by populations of elements. Furthermore, it is possible to conceive of a version of C-AEM in which the stimuli and configurational cues are represented by populations of elements, a different random sample of which are activated each time a stimulus or compound is presented.

Second, in C-AEM the activation level of a unit affects both the unit's contribution to the net associative strength of a pattern (Equation 13), and the change in associative strength of the unit following a trial (Equation 12). In REM an element is either active or inactive; if it is active then its associative strength may be modified according to Equation 1. Populations of elements will, therefore, accrue different total amounts of associative strength depending on their size, but each element will contribute its full associative strength to the prediction of the US if it is active. Given that Glautier (2007) has shown that REM may also be simulated by representing each population as a single binary unit, and scaling changes in association strength, the differences in the predictions of REM and C-AEM are not likely to be due to either the effects of unit activation on net associative strength or on changes in associative strength, but rather the combination of these two factors.

The third difference between the models is, however, especially important for the complex negative patterning and parity discrimination tasks; in C-AEM a stimulus does not activate units which explicitly signal the absence of other stimuli, whereas the inhibited elements in REM serve this

purpose. When a stimulus is presented alone, in C-AEM it will activate its own representational unit fully, and no others. That means that in a complex negative patterning or parity discrimination where A, B, and C are individually paired with the US, units *a, b,* and *c* will each have an asymptotic associative strength of λ. This is not affected by the value of *r*. In REM, however, this associative strength is distributed between four populations of elements (context-independent A_i, and context-dependent elements $A_{\sim b}$, $A_{\sim c}$, and $A_{\sim b \sim c}$), and the distribution of associative strength between these populations is affected by *r*. The result is a complex interplay between the expression of this associative strength on compound conditioning trials and changes in the associative strength of configurational cues (or context-dependent elements) dependent on *r*. The effect is significant variation in the distribution of associative strength between units or populations of elements between the two models.

In conclusion, I have presented a context-dependent added elements model which is a version of the RW model with configurational cues in which the activation of units representing stimuli and their configurations are dependent upon the context in which a stimulus is presented. C-AEM can accommodate a wide range of conflicting data from Pearce's and Wagner's laboratories concerning the effects of stimulus similarity on generalization and discrimination learning if we assume that the context-dependency of representational units is affected by the perceptual properties of stimuli. The success of C-AEM relative to both REM and Configural Theory suggest that a configurational cue approach put forward by Spence (1952) and adopted by Wagner and Rescorla (1972) still has something to offer contemporary models of associative learning.

References

Andrews, G. (1990). Euler's "exemplum memorabile inductionis fallacies" and q-Trinomial Coefficients. *Journal of the American Mathematical Society, 3,* 653-669.

Andrews, G., & Baxter, R. J. (1987). Lattice gas generalization of the hard hexagon model. III. q-Trinomial Coefficients. *Journal of Statistical Physics, 47,* 297-330.

Aydin, A., & Pearce, J. M. (1995). Summation in autoshaping with short- and long-duration stimuli. *Quarterly Journal of Experimental Psychology, 48B*, 215-234.

Aydin, A., & Pearce, J. M. (1997). Some determinants of response summation. *Animal Learning & Behavior, 25*, 108-121.

Atkinson, R. C., & Estes, W. K. (1963). Stimulus sampling theory. In R. D. Luce, R. R. Bush, & E. Galanter. (Eds.), *Handbook of Mathematical Psychology*. Volume 2. New York: John Wiley & Sons.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*, 115-147.

Bouton, M. E., Doyle-Burr, C., & Vurbic, D. (2012). Asymmetrical generalization of conditioning and extinction from compound to element and element to compound. *Journal of Experimental Psychology: Animal Behavior Processes, 38*, 381-393.

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

Brandon, S. E., & Wagner, A. R. (1998). Occasion setting: Influence of conditioned emotional responses and configural cues. In N. Schmajuk & P. C. Holland (Eds.), *Occasion setting: Associative learning and cognition in animals* (pp. 343-382). Washington, DC: American Psychological Association.

Bush, R. R., & Mosteller, F. (1955). *Stochastic Models for Learning*. New York: John Wiley & Sons.

Estes, W. K. (1950). Towards a statistical theory of learning. *Psychological Review, 57,* 94-107.

Estes, W. K. (1955a). Statistical theory of spontaneous recovery and regression. *Psychological Review, 62*, 145-154.

Estes, W. K. (1955b). Statistical theory of distributional phenomena in learning. *Psychological Review, 62*, 369-377.

George, D. N. (2018). Stimulus similarity affects patterning discrimination learning. *Journal of Experimental Psychology: Animal Learning & Cognition, 44*, 128-148.

Ghirlanda, S. (2015). On elemental and configural models of associative learning. *Journal of Mathematical Psychology, 64-65*, 8-16.

Glautier, S. (2004). Asymmetry of generalization in causal learning. *Quarterly Journal of Experimental Psychology, 57B*, 315-329.

González, F., Quinn, J. J., & Fanselow, M. A. (2003). Differential effects of adding and removing components of a context on the generalization of conditional freezing. *Journal of Experimental Psychology: Animal Behavior Processes, 29*, 78-83.

Harris, J. A. (2006). Elemental representations of stimuli in associative learning. *Psychological Review, 113*, 584-605.

Harris, J. & Livesey, E. (2010). An Attention-Modulated Associative Network. *Learning & Behavior, 38*, 1-26.

Hendry, J. S. (1982). Summation of undetected excitation following extinction of the CER. *Animal Learning & Behavior, 10*, 476-482.

Hull, C. L. (1949). Stimulus intensity dynamism (*V*) and stimulus generalization. *Psychological Review, 56*, 67-76.

Kamin, L. J. (1969). Selective association and conditioning. In N. J. Mackintosh & W. K. Honig (Eds.), *Fundamental issues in associative learning.* Halifax, Canada: Dalhousie University Press.

Kehoe, E. J. (1986). Summation and configuration in conditioning of the rabbit's nictitating membrane response to compound stimuli. *Journal of Experimental Psychology: Animal Behavior Processes, 12*, 186-195.

Kehoe, E. J., Horne, A. J., Horne, P. S., & Macrae, M. (1994). Summation and configuration between and within sensory modalities in classical conditioning in the rabbit. *Animal Learning & Behavior, 22*, 19-26.

Kinder, A., & Lachnit, H. (2003). Similarity and discrimination in human Pavlovian conditioning. *Psychophysiology, 40*, 226-234.

Kokkola, N. H., Mondragón, E., & Alonso, E. (2019). A double error dynamic asymptote model of associative learning. *Psychological Review, 126*, 506-549.

Konorski, J. (1948). *Conditioned Reflexes and Neuron Organization*. Cambridge: Cambridge University Press.

Livesey, E. J., Thorwart, A., & Harris, J. A. (2011). Comparing positive and negative patterning in human learning. *Quarterly Journal of Experimental Psychology, 64*, 2316-2333.

Myers, K. M., Vogel, E. H., Shin, J., & Wagner, A. R. (2001). A comparison of the Rescorla-Wagner and Pearce models in a negative patterning and a summation problem. *Animal Learning & Behavior, 29*, 36-45.

McLaren, I. P. L., Kaye, H., & Mackintosh, N. J. (1989). An associative theory of the representation of stimuli: Applications to perceptual learning and latent inhibition. In R. G. M. Morris (Ed.), *Parallel distributed processing - Implications for psychology and neurobiology*. Oxford, UK: Oxford University Press.

McLaren, I. P. L., & Mackintosh, N. J. (2000). An elemental model of associative learning: I. Latent inhibition and perceptual learning. *Animal Learning & Behavior, 28*, 211-246.

McLaren, I. P. L., & Mackintosh, N. J. (2002). Associative learning and elemental representation: II. Generalization and discrimination. *Animal Learning & Behavior, 30*, 177-200.

Miles, C. G., & Jenkins, H. M. (1973). Overshadowing in operant conditioning as a function of discriminability. *Learning and Motivation, 4*, 11-27.

Pavlov, I. P. (1927). *Conditioned reflexes.* London: Oxford University Press.

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

Pearce, J. M. (1994). Similarity and discrimination: A selective review and a connectionist model. *Psychological Review, 101,* 587-607.

Pearce, J. M. (2002). Evaluation and development of a connectionist theory of configural learning. *Animal Learning and Behavior, 30*, 73-95.

Pearce, J., M., Adam, J., Wilson, P. N., & Darby, R. J. (1992). Effects of discrimination training on responding during a compound conditioned stimulus. *Journal of Experimental Psychology: Animal Behavior Processes, 18*, 379-386.

Pearce, J. M., Aydin, A. and Redhead, E. S. (1997). Configural analysis of summation in autoshaping. *Journal of Experimental Psychology: Animal Behavior Processes, 23*, 84-94.

Pearce, J. M., Esber, G. R., George, D. N., & Haselgrove, M. (2008). The nature of discrimination learning in pigeons. *Learning and Behavior, 36*, 188-199.

Pearce, J. M., & George, D. N. (2002). The effects of using stimuli from three different dimensions on autoshaping with a complex negative patterning discrimination. *Quarterly Journal of Psychology, 55B*, 349-364.

Pearce, J. M., George, D. N., & Aydin, A. (2002). Summation: Further assessment of a configural theory. *Quarterly Journal of Experimental Psychology, 55B*, 61-73.

Pearce, J. M., George, D. N., Redhead, E. S., Aydin, A., & Wynne, C. (1999). The influence of background stimuli on summation in autoshaping. *Quarterly Journal of Experimental Psychology, 52B*, 53-74.

Pearce, J. M., & Redhead, E. S. (1993). The influence of an irrelevant stimulus on two discriminations. *Journal of Experimental psychology: Animal Behavior Processes, 19*, 180-190.

Pearce, J. M., Redhead, E. S., & George, D. N. (2002). Summation in autoshaping is affected by the similarity of the visual stimuli to the stimulation they replace. *Journal of Experimental Psychology: Animal Behavior Processes, 28*, 175-189.

Qadri, M. A. J., & Cook, R. G. (2015). Experimental divergences in the visual cognition of birds and mammals. *Comparative Cognition and Bheavior Reviews, 10*, 73-105.

Redhead, E. S. (2007). Multi-modal discrimination learning in humans: evidence for configural theory. *Quarterly Journal of Experimental Psychology, 60*, 1477-1495.

Redhead, E. S., & Curtis, C. (2013). Common elements enhance or retard negative patterning discrimination learning depending on modality of stimuli. *Learning and Motivation, 44*, 46–59.

Redhead, E. S., & Pearce, J. M. (1995). Similarity and discrimination learning. *Quarterly Journal of Experimental Psychology, 48B*, 46-66.

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

Rescorla, R. A. (1999). Associative changes in elements and compounds when the other is reinforced. *Journal of Experimental Psychology: Animal Behavior Processes, 25*, 247-255.

Rescorla, R. A. (2000). Associative changes in excitors and inhibitors differ when they are conditioned in compound. *Journal of Experimental Psychology: Animal Behavior Processes, 26*, 428-438

Rescorla, R. A., & Coldwell, S. E. (1995). Summation in autoshaping. *Animal Learning and Behavior, 23*, 314–326.

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 theory and research (pp. 64-99). New York: Appleton-Century-Crofts.

Rilling, M., De Marse, T., & La Claire, Luke. (1993). Contour deletion as a method for identifying the weights of features underlying object recognition. *Quarterly Journal of Experimental Psychology, 46B*, 43-61.

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

Spence, K. W. (1936). The nature of discrimination learning in animals. *Psychological Review, 43*, 427-449.

Spence, K. W. (1937). The differential response in animals to stimuli varying within a single dimension. *Psychological Review, 44*, 430-444.

Spence, K. W. (1952). The nature of the response in discrimination learning. *Psychological Review, 59*, 89-93.

Thein, T., Westbrook, R. F., & Harris, J. A. (2008). How the associative strengths of stimuli combine in compound: summation and overshadowing. *Journal of Experimental Psychology: Animal Behavior Processes, 34*, 155-166.

Thorwart, A. & Lachnit, H. (2009). Symmetrical generalization decrements: Configural stimulus processing in human contingency learning. *Learning & Behavior, 37*, 107-115.

Thorwart, A. & Lachnit, H. (2010). Generalization decrements: Further support for flexibility in stimulus processing. *Learning & Behavior, 38*, 367-373.

Thorwart, A., & Lachnit, H. (2020). Inhibited Elements Model – Implementation of an associative learning theory. *Journal of Mathematical Psychology, 94,* 102310.

Thorwart, A., Livesey, E. & Harris, J. (2012). Normalization between stimulus elements in a model of Pavlovian conditioning: Showjumping on an elemental horse. *Learning & Behavior, 40*, 334-346.

Thorwart, A., Uengoer, M., Livesey, E. J., & Harris, J. A. (2017). Summation effects in human learning: evidence from patterning discrimination in goal tracking. *Quarterly Journal of Experimental Psychology, 70*, 1366-1379.

Wagner, A. R. (1971). Elementary associations. In H. H. Kendler & J. T. Spence (Eds.), Essays in neobehaviorism: A memorial volume to Kenneth W. Spence (pp. 187-213). New York: Appleton-Century-Crofts.

Wagner, A. R. (2003). Context-sensitive elemental theory. *Quarterly Journal of Experimental Psychology, 23B*, 7-29.

Wagner, A. R., & Brandon, S. E. (2001). A componential theory of Pavlovian conditioning. In R. Mowrer & S. Klein (Eds.), *Handbook of contemporary learning theories* (pp. 23-63). Mahwah, NJ: Erlbaum.

Wagner, A. R., Logan, F. A., Haberlandt, K., & Price, T. (1968). Stimulus selection in animal discrimination learning. *Journal of Experimental Psychology, 76*, 171–180.

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: Academic Press.

Wheeler, D. S., Amundson, J. C., & Miller, R. R. (2006). Generalization decrement in human contingency learning. *Quarterly Journal of Experimental Psychology, 59*, 1212-1223.

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

Footnotes

1. Predictions derived from the RW model in this case are relatively insensitive to variations in the absolute values of the α and β parameters, as long as the α of all of the stimuli is the same, and the α of all of the configurational cues is the same. In general, whether the β associated with the presentation of the US is greater, less than, or the same as the β associated with the absence of the US has no impact on the model's predictions. Changes in the relative salience of the stimuli and configurational cues result in small differences in their asymptotic associative strengths, but unless extreme values are selected do not prevent the model from solving the discrimination. If all stimuli and configurational cues have the same α, the RW model predicts that the asymptotic associative strengths of the stimuli and configurational cues will be: A/B/C/D = 0.2λ, w/x = 0.6λ, y/z = –0.4λ.
2. It is common to assume that the value of β is greater when the US is present than when it is absent. Some effects are only predicted by RW model when this is the case. One example is the relative validity effect (Wagner, Logan, Haberlandt & Price, 1968) where less conditioned responding to stimulus X was observed following AX+ BXø training than following training in which compounds AX and BX were each reinforced on 50% of trials. When the two β values are equal the RW model predicts no difference between conditions. All simulations reported here were repeated with β equal to .05 both when the US was present and when it was absent. While this change had a trivial effect on the rate at which learning progressed, it did not alter the overall pattern of results for any simulation.

Author note

Correspondence concerning this article should be addressed to David. N. George, Department of Psychology, University of Hull, Hull HU67RX, UK. E-mail: d.george@hull.ac.uk. The author would like to thank Bill Whitlow and two anonymous reviewers for the helpful comments they made about an earlier version of this paper. The MATLAB code used to conduct the computer simulations reported here is available from the 'learnSim' repository on GitHub (GitHub Inc., San Francisco, CA) at https://github.com/DavidNGeorge/learnSim.

Appendices

Appendix A: Derivation of population expansion in REM using trinomial theorem

We may represent the effect of adding a stimulus (B) on individual elements within the representation of another stimulus (A) using balanced ternary notation where the element is either context dependently activated (+1; i.e. A_b), content-independent (0; i.e. A_i), or context dependently inhibited (-1, i.e. $A_{\sim b}$). Since replacement by different stimuli is statistically independent, we can then characterize the effect of any number of different stimuli upon a particular population of elements within the representation of A using a vector of ternary bits. In any system where *n* stimuli may be presented in combination, the vectors describing the various populations of elements within the representation of each stimulus will consist of *n – 1 = m* bits. For example, within the representation of A, elements that are context-dependently inhibited by B, but are context-independent with respect to C ($A_{\sim b}$) may be represented by the vector [-1, 0], and elements that are context-dependently activated only in the presence of both B and C (A_{bc}) would be represented by the vector [+1, +1]. The number of populations for which these vectors sum to a particular value, *k*, is given by the trinomial coefficient which may be calculated using Equation A1.

$$\binom{m}{k}_2 = \sum_{j=0}^{m} \frac{m!}{j!(j+k)!(m-2j-k)!} \qquad \text{(A1)}$$

where

$$\binom{m}{-k}_2 = \binom{m}{k}_2 \qquad \text{(A2)}$$

(Andrews 1990; Andrews & Baxter, 1987) and *k* has the range (*-m, m*). The total number of populations of elements *within the representation of A*, is then given by the term $\sum_{k=-m}^{m} \binom{m}{k}_2$. Since the trinomial coefficients are defined by Equation A3 and *m = n* – 1*,* if we substitute 1 for *x* in Equation A3, we can see that it reduces to 3^{n-1}. This is the number of populations within the representation of a

single stimulus. For a system consisting of *n* stimuli, $n3^{n-1}$ distinct populations of elements may, therefore, be identified.

$$\left(1+x+x^{-1}\right)^{m}=\sum_{k=-m}^{m}\binom{m}{k}_{2}x^{k} \tag{A3}$$

Appendix B: General form of the C-AEM activation function

Equation 9 gives the activation function for units in C-AEM when all *r* values and stimulus intensities are equal. In many situations this will, however, not be the case. For example, when a stimulus may be presented in compound with other stimuli from either the same or a different modality. A general form of the activation function is given in Equation B1 which allows stimuli to differ both in their intensity and the extent to which they interact with each other. In this equation, I_H is the intensity of stimulus H, $r_{j,h}$ is the extent to which stimulus J interacts with the perception of stimulus H, *n* is the total number of stimuli present and *k* is the number of stimuli that contribute to the activation of the unit.

$$\gamma = \sum_{h=1}^{k} I_H \prod_{i=1, i\neq h}^{k} r_{i,h} \prod_{j=k+1}^{n} (1 - r_{j,h}) \tag{B1}$$

In a stimulus compound, each stimulus will contribute to the activation of more than one unit. When there are three or more stimuli present, then each stimulus will contribute to the activation of multiple configurational units. Because of this, activation of a unit may be affected by the presence of stimuli which do not directly contribute to its activation. Consider compound ABC. Stimuli A will contribute to the activation of units *a*, *ab*, *ac*, and *abc*. Because stimulus A always contributes the same total activity to the system, equal to its intensity I_A, the presence of stimulus C will reduce the amount of activity available for units *a* and *ab*. In Equation B1, the contribution of each stimulus (*h* = 1 to *k*) to the activation of a unit is calculated based on its intensity (I_H) and these contributions are summed (stimuli A and B both contribute directly to the activation of unit *ab*, all three stimuli contribute directly to the activation of unit *abc*). The intensity of each stimulus is multiplied by the product of the *r* values for each additional stimulus that contributes to the activation of the unit, and the product of the (1 – *r*) values for each additional stimulus that does not contribute to the activation of the unit. Hence, the contribution of stimulus A to the activation of the *ab* unit is $I_A r_{b,a}(1 - r_{c,a})$, and

the total activation of the unit, $\gamma_{ab} = I_A r_{b,a}(1 - r_{c,a}) + I_B r_{a,b}(1 - r_{c,b})$. Activation of unit *a* is $\gamma_a = I_A(1 - r_{b,a})(1 - r_{c,a})$, and of unit *abc* is $\gamma_{abc} = I_A r_{b,a} r_{c,a} + I_B r_{a,b} r_{c,b} + I_C r_{a,c} r_{b,c}$.

The total activity across all units cannot now be calculated using Equations 10 and 11. It will, however, always be equal to Σ*I*. It seems somewhat implausible to suggest that when two stimuli of greatly differing intensities are presented in compound that the less intense stimulus will interfere with the more intense stimulus to the same extent as the more intense stimulus interferes with the less intense. This may be illustrated with an example in which stimulus A has an intensity of 4, and stimulus B an intensity of 1. If the replacement parameter, $r_{b,a}$, has a value of 0.4, then from Equation B1 we can see that the activation levels of *a, b*, and *ab*, will be 2.4, 0.6, and 2, respectively. Hence, the configurational cue *ab* is of greater intensity than one of the stimuli contributing to its activation when presented alone. For this reason, I propose that two further modifications might be appropriate. First, perceptual interaction between two cues may not always be symmetrical: the degree to which B interferes with A, $r_{b,a}$, is not necessarily equal to the degree to which A interferes with B, $r_{a,b}$. Second, the value of the parameter $r_{b,a}$ may be proportional to the ratio between the intensities of the two stimuli A and B: the greater the ratio A:B, the greater value of $r_{a,b}$ and the smaller the value of $r_{b,a}$. These modifications do not affect the total contribution that each stimulus makes to activity across all units; γ_{total} will always be equal to Σ*I*. Exploration of the effects of these modifications are, however, beyond the scope of this article which considers only situations in which all stimulus intensities and *r* values are equal.

Appendix C: Derivation of population expansion in C-AEM using binomial theorem

C-AEM exhibits a similar exponential growth in the number of units required to represent stimulus compounds as does REM, but at a much slower rate. The total number of units that may be activated in a system consisting of *n* stimuli is equal to the number of units activated by the compound containing all *n* stimuli: one unit for each of the *n* stimuli, plus units activated by each combination of 2 or more stimuli in the set. Since the number of units receiving activation from *k* inputs is given by the binomial coefficient (see Equation 8 in the main text), then the total number of units activated by an *n* stimulus compound must be $\sum_{k=1}^{n}\binom{n}{k}$. Binomial coefficients may be defined by Equation C1.

$$(1+x)^n = \sum_{k=0}^{n}\binom{n}{k}x^k \tag{C1}$$

If we substitute 1 for *x* in Equation C1, we can see that it reduces to 2^n. This total includes the case where *k* = 0, but we do not require a unit to which no stimuli contribute activation. Since $\binom{n}{0}=1$, the total number of units required to represent a system in which *n* stimuli may be presented in compound is, therefore, $(2^n - 1)$.

Tables

Table 1. The populations of elements within the representation of a stimulus, A, according to Wagner's REM. There are nine distinct populations to consider when stimulus A may be presented in combination with either or both of two other stimuli, B and C. Each population of elements may be context-independent, context-dependently activated (added), or context-dependently inhibited with respect to B and with respect to C. The relative size of the populations is determined by parameters r_b and r_c. Whenever A is presented, whether alone or in compound with B and/or C, four of these populations of elements will be active. Activity within those populations will always sum to 1.

Dependence on Stimulus C	Dependence on Stimulus B: Independent	Dependence on Stimulus B: Presence	Dependence on Stimulus B: Absence
Independent	A_i $(1 - r_b)(1 - r_c)$	A_b $r_b(1 - r_c)$	$A_{\sim b}$ $r_b(1 - r_c)$
Presence	A_c $(1 - r_b)r_c$	A_{bc} $r_b r_c$	$A_{c\sim b}$ $r_b r_c$
Absence	$A_{\sim c}$ $(1 - r_b)r_c$	$A_{b\sim c}$ $r_b r_c$	$A_{\sim b\sim c}$ $r_b r_c$

Figure Captions

Figure 1. The predictions concerning the relative size of the effects of overshadowing and external inhibition for Pearce's (1994) Configural Theory, Wagner's (2003) Replacement Elements Model (REM), and the Context-dependent Added Elements Model (C-AEM). Top-left panel: generalization between patterns A and AB when either A is presented at test following conditioning with AB (remove), or AB is presented at test following conditioning with A (add); the predictions of Configural Theory are the same in both cases. The other three panels show the results of corresponding tests of generalization between patterns A and ABC (top-right), AB and ABC (bottom-left) and ABC and ABCD (bottom-right).

Figure 2. The predicted results of a test of simple summation in which two stimuli are presented either alone (A/B) or in compound (AB) following conditioning to the individual stimuli (A+ B+). Top panel: Pearce's (1994) Configural Theory without (left) or with (right) the inclusion of a salient contextual cue (X) throughout conditioning and at test. Bottom panel: Wagner's (2003) Replaced Elements Model (REM) and the Context-dependent Added Element Model (C-AEM) as a function of the replacement/perceptual interaction parameter, *r*.

Figure 3. The predicted results of a test of differential summation in which compound ABC is presented following conditioning to the individual stimuli (A+ B+ C+) or with pairs of stimuli (AB+ AC+ BC+). Top panel: Pearce's (1994) Configural Theory without (left) or with (right) the inclusion of a salient contextual cue (X) throughout conditioning and at test. Middle panel: Wagner's (2003) Replacement Elements Model (REM) as a function of the perceptual interaction parameter, *r*. Bottom panel: The Context-dependent Added Element Model (C-AEM) as a function of the perceptual interaction parameter, *r*.

Figure 4. Predictions concerning the course of acquisition of a complex negative patterning discrimination (A+ B+ C+ AB+ AC+ BC+ ABCø). Top row: Pearce's (1994) Configural Theory either without (left panel) or with (centre panel) the inclusion of a salient contextual cue. Middle row:

Wagner's (2003) Replaced Elements Model (REM) with either a low (left panel), intermediate (centre panel) or a high (right panel) value of the replacement parameter, *r*. Bottom row: The Context-dependent Added Elements Model (C-AEM) with either a low (left panel), intermediate (centre panel) or a high (right panel) value of the perceptual interaction parameter, *r*.

Figure 5. Predictions concerning the course of acquisition of a parity discrimination (A+ B+ C+ ABø ACø BCø ABC+). Top row: Pearce's (1994) Configural Theory with different values of the discriminability parameter, *d* (left panel, *d* = 2; centre panel, *d* = 3). Middle row: Wagner's (2003) Replaced Elements Model (REM) with either a low (left panel), intermediate (centre panel) or a high (right panel) value of the replacement parameter, *r*. Bottom row: The Context-dependent Added Elements Model (C-AEM) with either a low (left panel), intermediate (centre panel) or a high (right panel) value of the perceptual interaction parameter, *r*. Note that the scale for the bottom-left panel is an order of magnitude different from the other panels in the middle and bottom rows.

Figure 6. Predictions concerning the course of acquisition of a negative (A+ B+ C+ ABø ACø BCø ABC+) parity discriminations with symmetrical outcomes. Top row: Pearce's (1994) Configural Theory with different values of the discriminability parameter, *d* (left panel, *d* = 2; centre panel, *d* = 3). Middle row: Wagner's (2003) Replaced Elements Model (REM) with either a low (left panel), intermediate (centre panel) or a high (right panel) value of the replacement parameter, *r*. Bottom row: The Context-dependent Added Elements Model (C-AEM) with either a low (left panel), intermediate (centre panel) or a high (right panel) value of the perceptual interaction parameter, *r*. Note that the scale for the bottom-left panel is an order of magnitude different from the other panels in the middle and bottom rows.

Figures

Figure 1

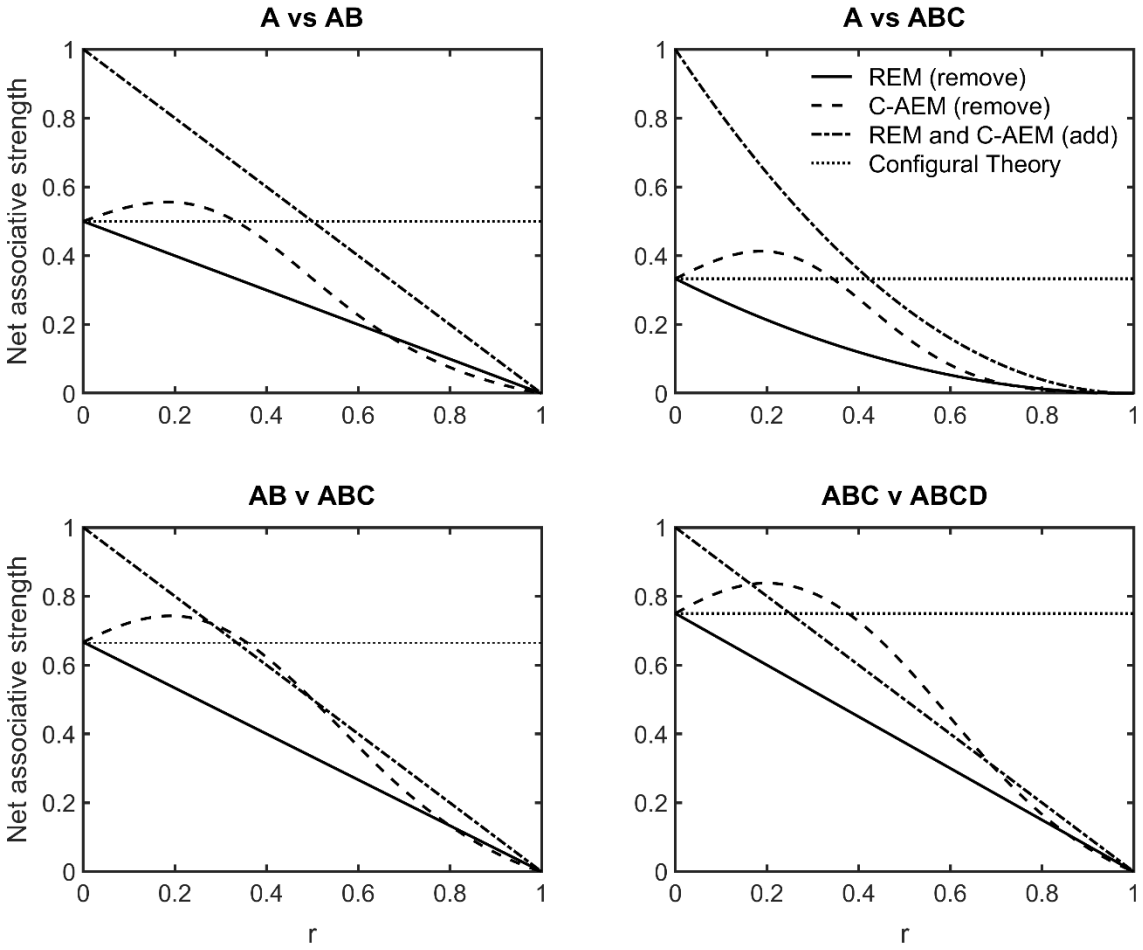


Figure 2

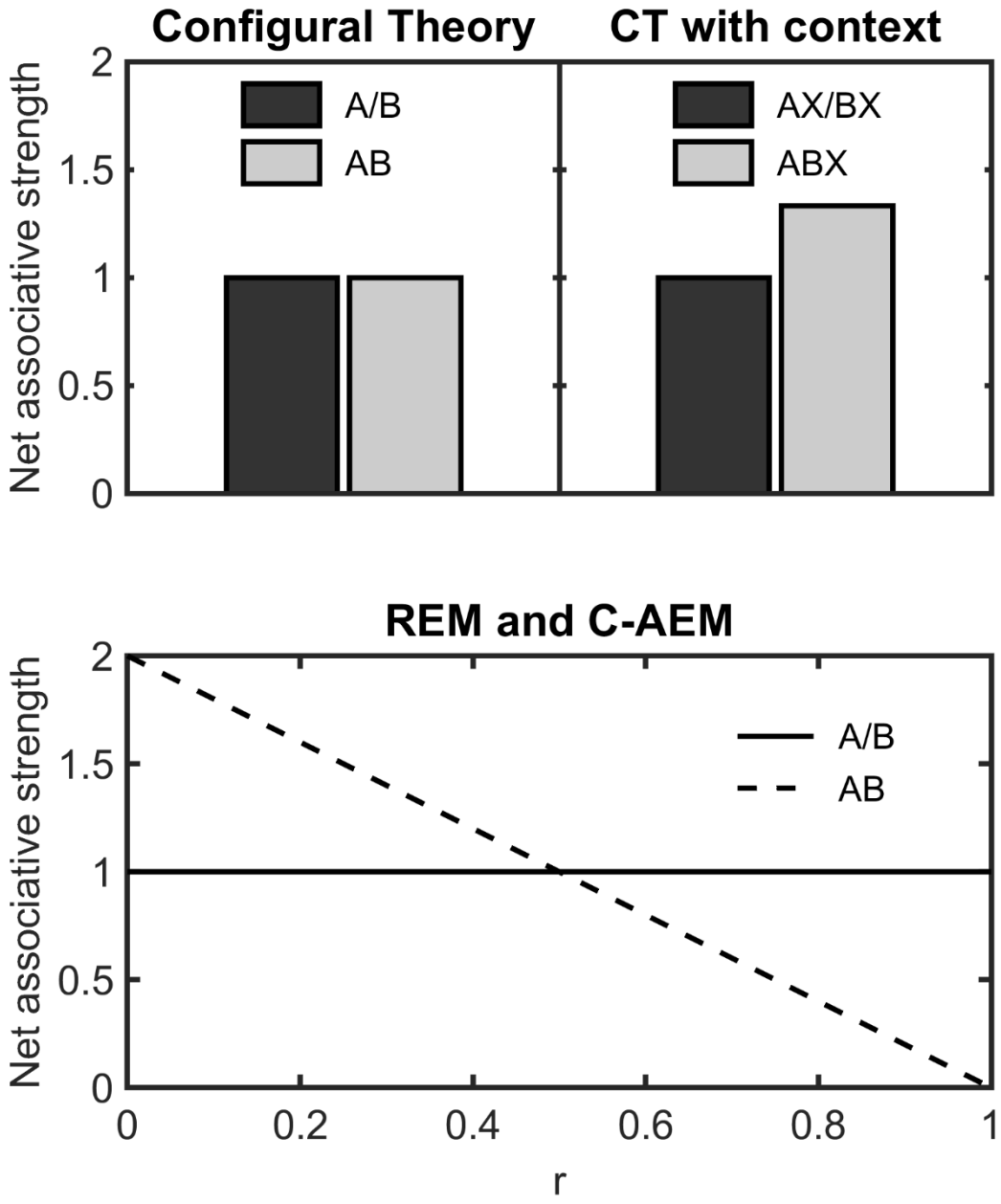

Configural Theory
CT with context
Net associative strength
2
1.5
1
0.5
0
A/B
AB
AX/BX
ABX
REM and C-AEM
Net associative strength
2
1.5
1
0.5
0
A/B
AB
0
0.2
0.4
0.6
0.8
1
r

Figure 3

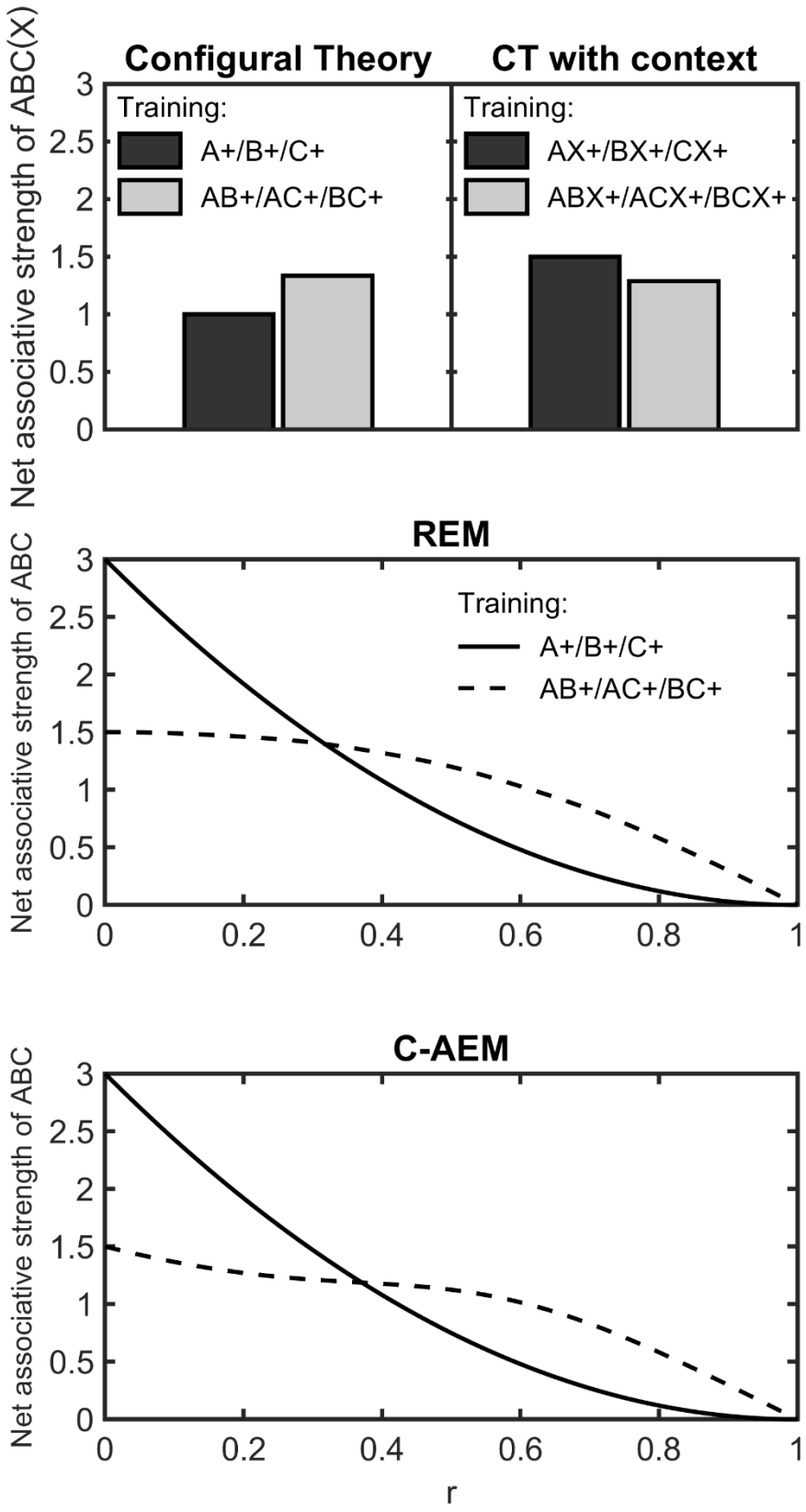

Configural Theory
CT with context
Training:
A+/B+/C+
AB+/AC+/BC+
Training:
AX+/BX+/CX+
ABX+/ACX+/BCX+
Net associative strength of ABC(X)
REM
Training:
A+/B+/C+
AB+/AC+/BC+
Net associative strength of ABC
C-AEM
Net associative strength of ABC
r

Figure 4

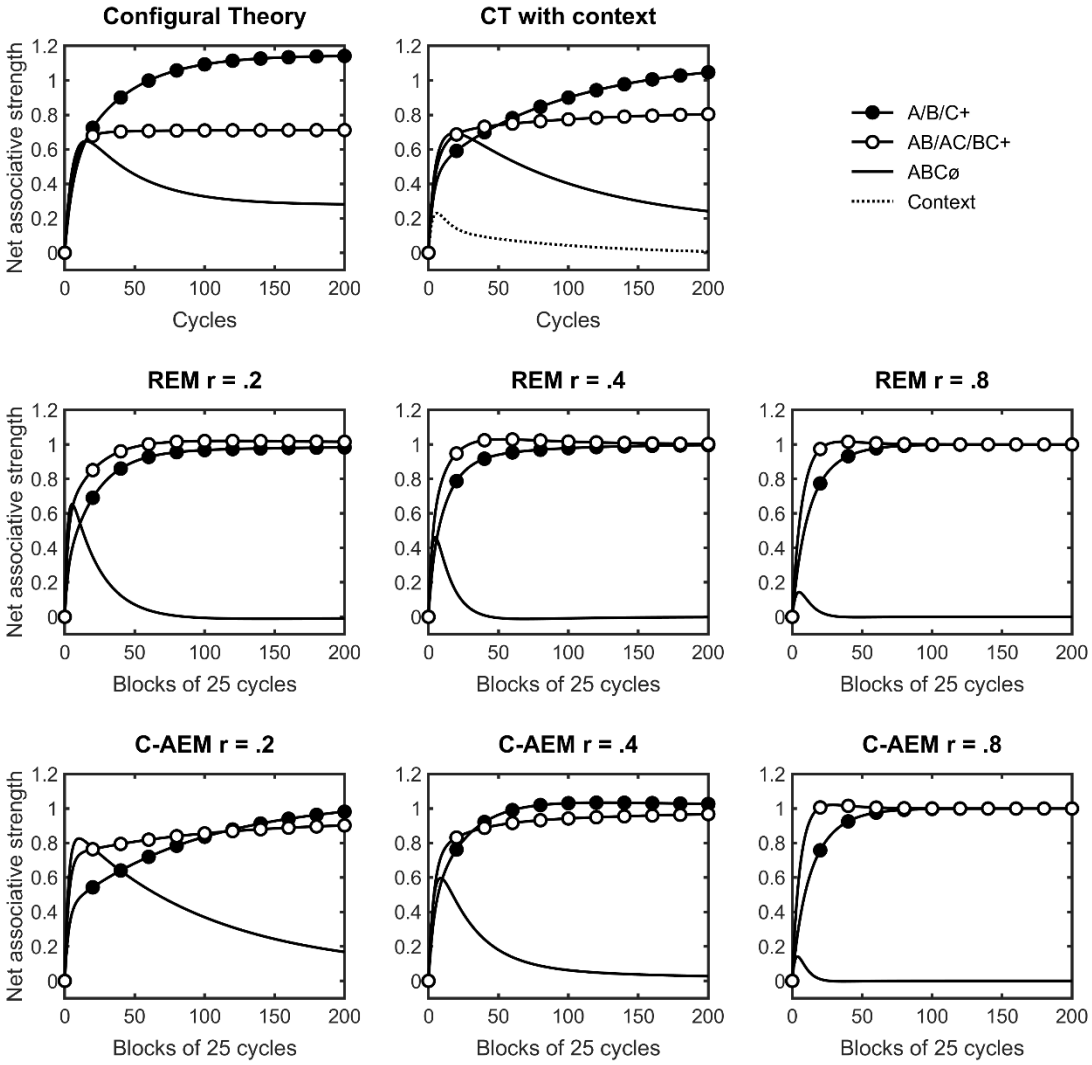

Configural Theory
CT with context
REM r = .2
REM r = .4
REM r = .8
C-AEM r = .2
C-AEM r = .4
C-AEM r = .8
Net associative strength
Cycles
Blocks of 25 cycles
A/B/C+
AB/AC/BC+
ABCø
Context

Figure 5

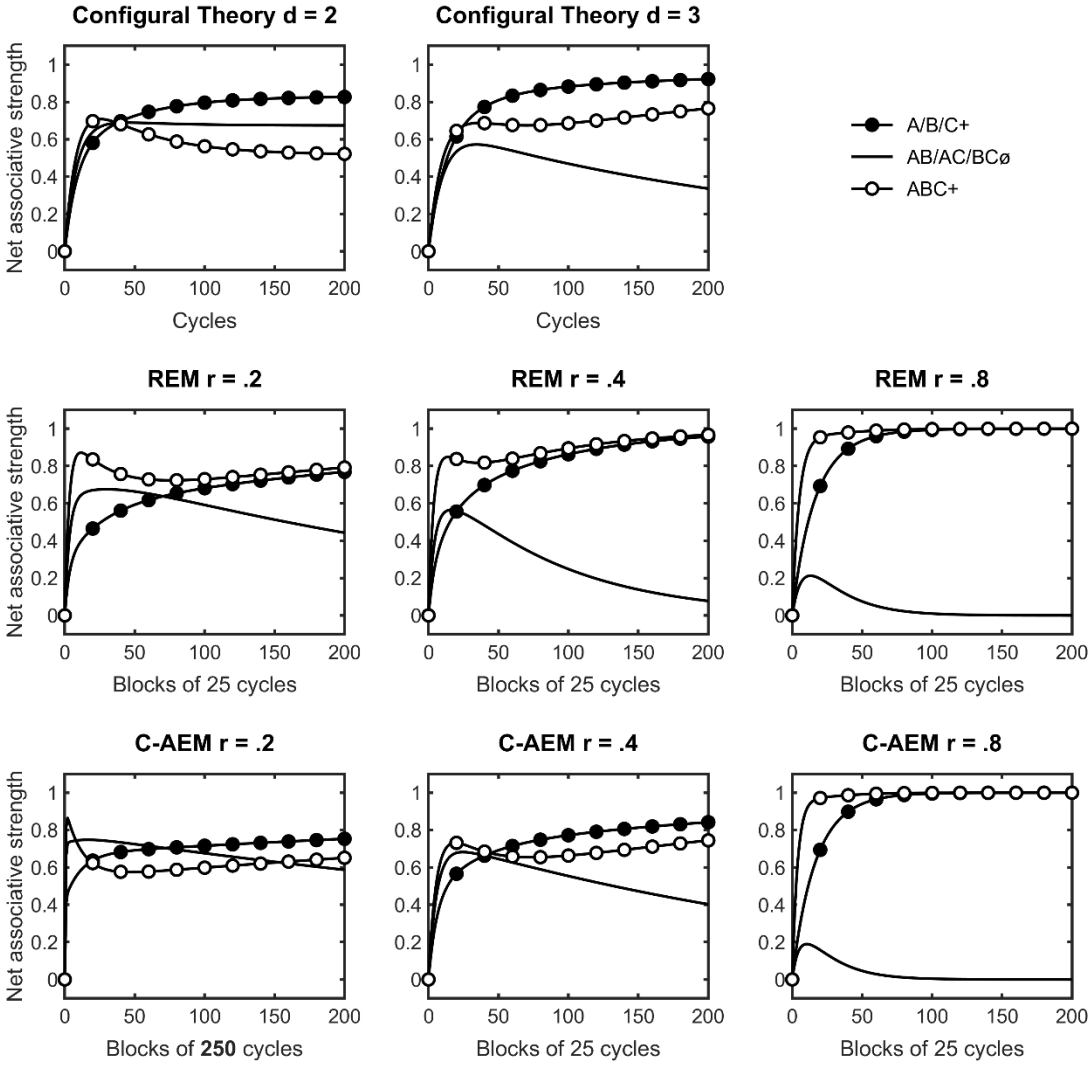
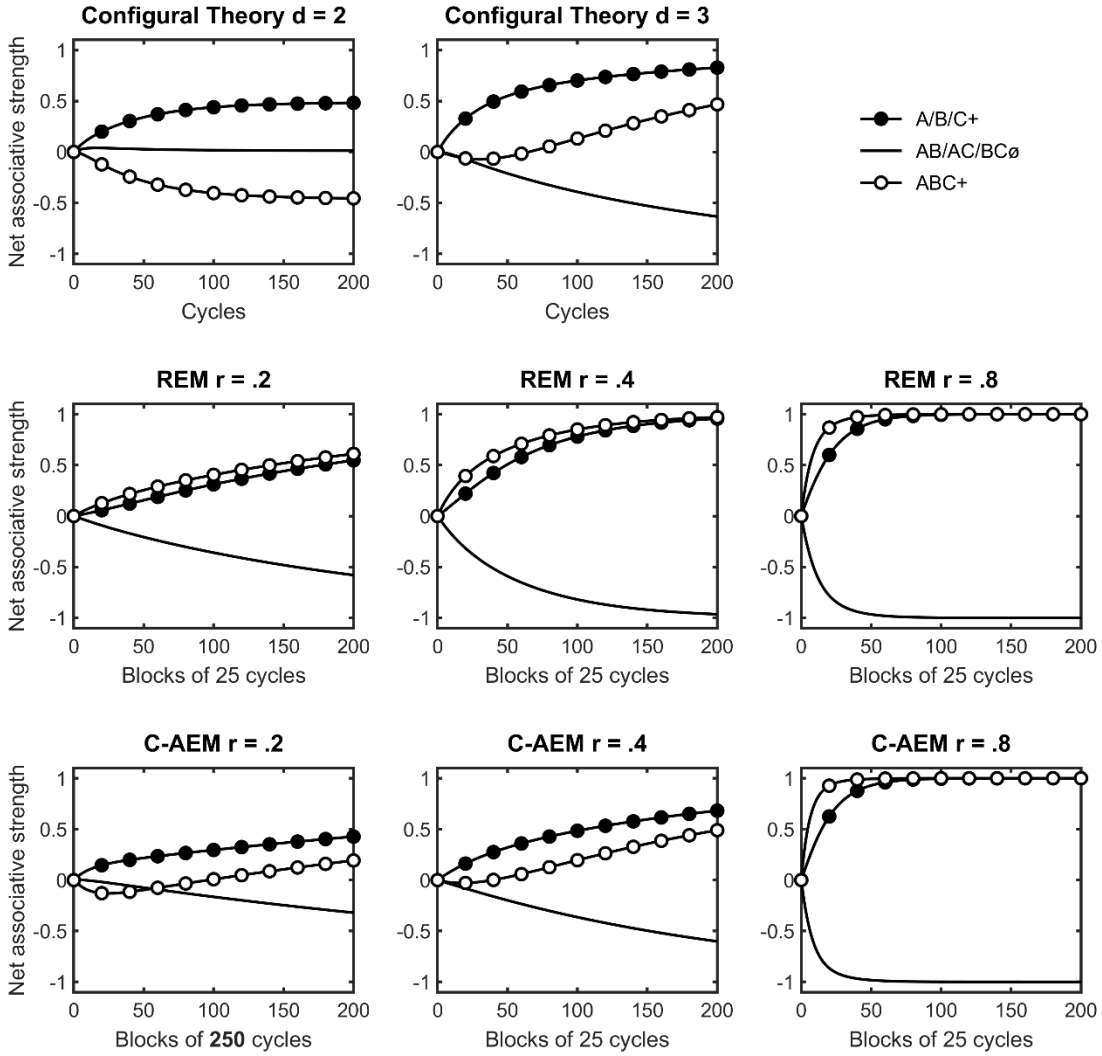

Configural Theory d = 2
Configural Theory d = 3
A/B/C+
AB/AC/BCø
ABC+
Net associative strength
Cycles
REM r = .2
REM r = .4
REM r = .8
Blocks of 25 cycles
C-AEM r = .2
C-AEM r = .4
C-AEM r = .8
Blocks of 250 cycles

Figure 6


Configural Theory d = 2
Configural Theory d = 3
A/B/C+
AB/AC/BCø
ABC+
Net associative strength
Cycles
REM r = .2
REM r = .4
REM r = .8
Blocks of 25 cycles
C-AEM r = .2
C-AEM r = .4
C-AEM r = .8
Blocks of 250 cycles