

Normalization between stimulus elements in a model of Pavlovian conditioning: Showjumping on an elemental horse

Anna Thorwart · Evan J. Livesey · Justin A. Harris

© Psychonomic Society, Inc. 2012

Abstract Harris and Livesey. *Learning & Behavior*, 38, 1–26, (2010) described an elemental model of associative learning that implements a simple learning rule that produces results equivalent to those proposed by Rescorla and Wagner (1972), and additionally modifies in “real time” the strength of the associative connections between elements. The novel feature of this model is that stimulus elements interact by suppressively normalizing one another’s activation. Because of the normalization process, element activity is a nonlinear function of sensory input strength, and the shape of the function changes depending on the number and saliences of all stimuli that are present. The model can solve a range of complex discriminations and account for related empirical findings that have been taken as evidence for configural learning processes. Here we evaluate the model’s performance against the host of conditioning phenomena that are outlined in the companion article, and we present a freely available computer program for use by other researchers to simulate the model’s behavior in a variety of conditioning paradigms.

Keywords Associative learning · Computational modelling

Most models of conditioning describe the content of learning as the strengthening or weakening of an association or link that connects some representations of the conditioned stimulus (CS) and the unconditioned stimulus (US). One fundamental issue that has been extensively debated concerns the distributed versus unitary nature of the way that stimuli are represented within the associative network. On one side of this distinction are theories that assume

individual stimuli to be represented by multiple elements distributed across the network; on the other side are theories in which whole stimulus patterns are represented by a single configural unit. Much of the debate has centered on how these models perform in solving particular “nonlinear” discriminations, such as negative patterning. The key to solving the discrimination is to provide a means by which some of what is learned on reinforced trials does not generalize to nonreinforced trials, or vice versa. Most viable models of associative learning (elemental or configural) achieve this by assuming that stimulus representations involve a nonlinear combination of stimulus elements (e.g., Pearce, 1994; Wagner 2008). In the case of purely elemental models (e.g., Harris, 2006; Harris & Livesey, 2010; McLaren & Mackintosh, 2000, 2002), there is *quantitative* nonlinearity in the processing of stimulus elements. For example, the elements representing a compound, AB, are the same as those representing the individual stimuli, A and B, but the strength of any element’s activation by the compound will often differ from its activation by A or B alone (or, if the element is common to A and B, its activation in the compound will differ from the simple sum of its activations by both stimuli). This means that most complex discriminations, including negative and positive patterning, biconditional discriminations, and some aspects of occasion setting, can be solved by relying solely on elemental representations.

The present article focuses on a model first proposed by Harris (2006), and subsequently elaborated by Harris and Livesey (2010). The model uses a simple normalization operation, attributed to a capacity limitation of attention, to regulate the activation strengths of the stimulus elements as a function of the total number of elements activated. We will give only a brief description of the model here, but we refer the reader to previous reports for detailed descriptions (Harris, 2006, 2010; Harris & Livesey, 2010). We will then go on to

A. Thorwart · E. J. Livesey · J. A. Harris (✉)
University of Sydney,
Sydney, New South Wales, Australia
e-mail: Justin.harris@sydney.edu.au

describe how the model performs when accounting for the variety of empirical phenomena that have been set out in the companion article, and we will present several illustrative simulations of the model. In doing this, we will concentrate on phenomena that are not already discussed at length in our previous articles.

An attention-modulated associative network (AMAN)

The model developed by Harris and Livesey (2010) considers stimuli to have distributed representations within an associative network of excitatory units (E). Each E unit is paired with an inhibitory unit (I) and an attention unit (A), and this triad is referred to in the following text as an *element* that corresponds to a certain feature of a stimulus (see Fig. 1). Within the network, each E has a fixed probability of being connected to the E unit of every other element (connectivity), and associative learning depends on changes in the strengths of these connections. The model treats CS and US elements as equivalent in terms of the rules governing their behavior and connectivity within the network, and associations within a single stimulus, between two CSs, or between a CS and a US are treated in the same manner. The activation strength of E units is subject to a gain control mechanism that normalizes the behavior of the network as a whole. In general terms, this means that the response of each E unit in the network to incoming stimulation is altered by the activity of other E units in the network. Normalization among Es is achieved through the network of corresponding I units. The activity of the I units is itself regulated by the network of paired A units, such that attention (driven by rising sensory input) reduces inhibition of E units by releasing them from the influence of their I units. Finally, the activation of each A unit is normalized by activity in other A units in the attention network. This normalization (gain control) creates a functional capacity limitation, such that the ability of a given sensory input to excite its attention unit is diminished by other sensory inputs that compete for attention.

Computations

Simulation of the model's behavior is achieved through a set of equations that determine the activation strength of each E, I, and A unit, on the basis of "external" sensory input (S) and "internal" excitatory and inhibitory inputs from other units in the network. At a given moment, each unit's response, R_{pot} , is a function of the unit's excitatory input, normalized by inhibitory inputs. A general form of this function for any E, I, or A unit can be specified as

$$R_{\text{pot}} = \frac{\text{Input}^p}{\text{Input}^p + N^p + D}, \quad (1)$$

where Input is the sum of the excitatory inputs to the unit, N is the sum of the normalizing inputs, and D is a constant that prevents the denominator from equaling zero.

Equation 1 derives from computational rules that have been used in numerous existing models of sensory systems that incorporate a gain control process of normalization (e.g., Grossberg, 1973; Heeger, 1992; Reynolds & Chelazzi, 2004; Reynolds & Heeger, 2009), and its application to the activation of elements has been discussed recently by Harris (2010). It is a monotonically increasing function that asymptotes at 1. If the inhibitory input $N = 0$, the excitatory inputs (Input^p) must equal the constant, D, to excite the unit to half of its potential response ($R = .5$). As N—that is, the normalization—increases, the function is shifted to the right, such that the strength of excitatory Input^p must increase by N^p in order for R to reach half height. The constant D scales the range of Input values over which critical changes in R occur, and the power p determines the slope of the function. When $p = 1$, the function is a simple monotonically increasing curve. Higher values of p make the function sigmoid, and increasing p increases the maximal slope.

Equation 1 defines the unit's response potential to its excitatory and inhibitory inputs at a given instant in time. However, we assume that units cannot change instantaneously from their existing activation level to the new level, since this would imply an infinite rate of change. Rather, the real response (R) of each unit gradually approaches the response potential (R_{pot}), as defined in Eq. 2:

$$\frac{dR}{dt} = \delta \times (R_{\text{pot}} - R). \quad (2)$$

Equation 1 describes the general form of the function relating the response of a unit to its input. Below, we give the specific equations for each E, I, and A unit. For E units, the excitatory inputs come from both sensory input S and other E units, and the normalizing inputs come from the I units; for I units, the excitatory inputs come from all E units and the normalizing inputs from A units; and for A units, the excitatory inputs come from S, whereas the normalizing inputs come from other A units (as illustrated in Fig. 1). We have set at zero the lower limit on the total summed input to any unit, constraining the activity in any unit to be nonnegative.

For the E unit of element X, the response of E_x is therefore

$$R_{\text{pot}}(E_x) = \frac{\text{Input}(E_x)^p}{\text{Input}(E_x)^p + I_x^p + D}, \quad (3)$$

where

$$\text{Input}(E_x) = S_x + \sum_{i=1}^n V_i \times E_i.$$

When calculating the activation of I units, their input is a weighted sum of activities in all E units, normalized by inhibition from A units. Thus, each I unit receives connections

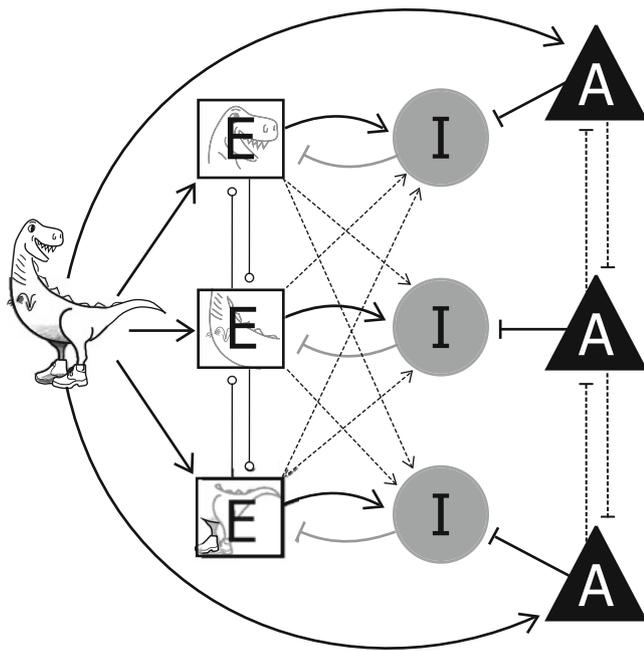


Fig. 1 The basic structure of the associative network described by Harris and Livesey (2010). The network is organized in triads of paired excitatory (E), inhibitory (I), and attention (A) units. Each triad is referred to as an *element*. Sensory input (a smiling dinosaur, in this example) activates an array of E units that represent different features of the stimulus (only three are shown here, for clarity). Each E activates a paired I and weakly activates (dashed arrows) the Is of other elements that have similar spatial and featural receptive fields. The I units inhibit their paired E unit, thereby normalizing the activation of each E depending on the extent of activation of the surrounding Es. Activity in I units is regulated by a network of attention units (A). Each A is excited by the same sensory input that activates the corresponding E, and that A unit inhibits the corresponding I unit, thus increasing activation of the E unit (releasing E from inhibition). Finally, inhibition among all A units (dashed lines) normalizes their activity. Learning involves changes in the strengths of connections between E units. For the sake of clarity, not all possible connections are displayed

from all Es, with each connection being weighted (z) in the range from 0 to 1. Larger values of $z_{i,x}$ reflect greater similarity between E_i and E_x , with the effect that E_x receives stronger suppressive normalization from those E_i units that have more similar sensory tuning. The input to each I unit is normalized by input from a specific A unit. This input is scaled by a factor, k_a , that allows the attention units to strongly suppress I units. Thus, the response of unit I_x is

$$R_{\text{pot}}(I_x) = \frac{\text{Input}(I_x)^p}{\text{Input}(I_x)^p + (k_a \times A_x)^p + D}, \tag{4}$$

where

$$\text{Input}(I_x) = \sum_{i=1}^n z_{i,x} \times E_i.$$

Activity in the attention units is driven by sensory input S , normalized by the activity (A') in every *other* A unit.

Thus, the attention field functions as a competitive network of fully connected units, where all connections are suppressive and have the same fixed strength (w):

$$R_{\text{pot}}(A_x) = \frac{S_x^p}{S_x^p + A_x'^p + D}, \tag{5}$$

where

$$A_x' = w \cdot \left(\left[\sum_{i=1}^n A_i \right] - A_x \right).$$

Associative change The Harris and Livesey (2010) model specifies associative processes in real time, using an algorithm in which associative change is determined by the activation state of the recipient element. In most respects, this algorithm functions in a fashion very similar to the summed error term of the Rescorla–Wagner (1972) model, because the instantaneous change in activation of the recipient element is generally proportional to the summed error term. The algorithm is similar to an idea first suggested by Konorski (1948) (see also McLaren & Dickinson, 1990; Sutton & Barto, 1981) in assuming that the strength of a connection between two Es increases when activity in the E unit of the recipient element rises. However, unlike those earlier proposals—which assumed that connection strength decreases when activity in the recipient E unit falls—the learning rule defined by Harris and Livesey (2010) states that the strength of the connection decreases when activity rises in the inhibitory (I) unit of the recipient element. Thus, rather than tying the direction of associative change to the direction of change in activity of the recipient excitatory E unit, the present rule ties the direction of associative change to the rise in excitation versus inhibition of the recipient element.

The change in V from the sending element X to recipient element Y , $\Delta V_{x,y}$, is defined in Eq. 6. It is *proportional* to the difference between the rise in E_y , $dR(E_y)/dt$, and the rise in I_y , $dR(I_y)/dt$, each scaled by its own rate parameter, β_E and β_I . This difference is then scaled by the activation strengths of the sending E unit (E_x) and the recipient E unit (E_y).¹ We set $\beta_E = 0$ if the change in E_y is negative, and likewise $\beta_I = 0$ if the change in I_y is negative, so that learning is unaffected by falls in activation.

$$\Delta V_{xy} = E_x \cdot E_y \cdot \left(\beta_E \cdot \frac{dR(E_y)}{dt} - \beta_I \cdot \frac{dR(I_y)}{dt} \right). \tag{6}$$

¹ Scaling $\Delta V_{x,y}$ by E_y was not included in the model as originally described by Harris and Livesey (2010), but is adopted here to balance the contributions of E_x and E_y on learning, and to prevent changes in $V_{x,y}$ when E_y has zero activation.

Simulating performance in experimental designs

We provide an easy-to-use simulator program developed to assist researchers in simulating the AMAN model proposed by Harris and Livesey (2010). The program, with accompanying instructions, can be freely downloaded from our website at <http://sydney.edu.au/science/psychology/staff/justinh/downloads/>.

The program is available as two .m files that can be run using MATLAB or as executable files that can be run as standalone programs, for use by researchers who do not have the MATLAB software package. In all simulations described in this article, the parameters were not changed, with the exception of the number of time steps per trial. As the simulations are computationally complex and time-consuming, the number of time steps was kept as small as possible. The full parameter file, as used in these simulations, is included in the Appendix of this article.

To demonstrate how the model performs on the list of conditioning phenomena outlined in the lead article of this volume, we have divided the to-be-explained phenomena into three categories. First, we will discuss those phenomena that are accounted for by the model in terms of a Rescorla–Wagner-like learning rule applied to elemental representations of stimuli. Second, we will discuss those phenomena that can be accounted for by suppressive normalization between elements and the operation of the attention mechanism, as well as the effects on element-to-element associations that develop across stimulus presentations. Finally, we will consider those phenomena that are beyond the explanatory scope of the model, at least in its current formulation.

Phenomena explained by a summed-error learning rule and elemental representations

Acquisition, extinction, and conditioned inhibition As described by Harris and Livesey (2010), the AMAN model describes both the acquisition and extinction of conditioned responding as gradual processes that are directly related to the accumulation (or loss) of associative strength. The model also predicts that the strength of conditioning will be lower for a partially reinforced (PRf) CS than for a continuously reinforced (CRf) one (Harris & Carpenter, 2011; Pavlov, 1927), because the associative strength of the PRf CS will extinguish during each nonreinforced trial. This difference in conditioning between CRf and PRf CSs is shown in a simulation presented in Fig. 2. The figure also shows simulations of extinction of responding to the two CSs when they are repeatedly presented without reinforcement in a subsequent phase; as is evident, the model does not predict slower extinction of the PRf CS (i.e., the partial-reinforcement extinction effect).

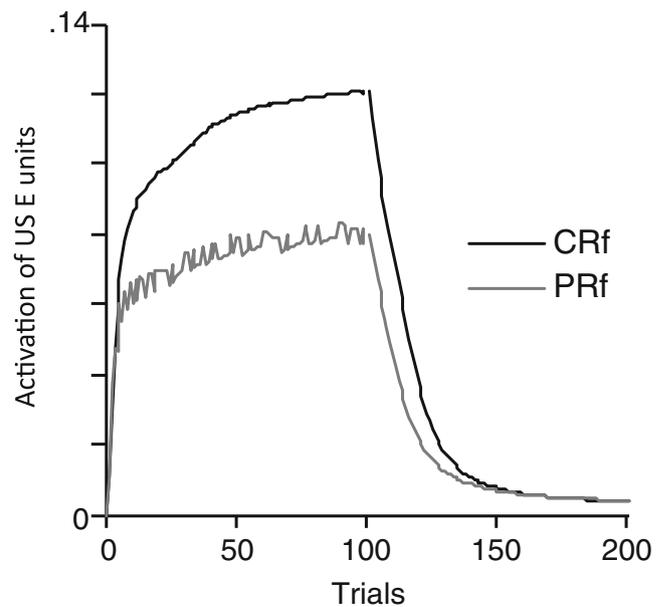


Fig. 2 Simulation of conditioned response strength (shown as the activation strength of US elements during the CS) to a continuously reinforced CS (CRf) and a partially reinforced (50 %) CS (PRf). The plot shows 100 trials of conditioning, followed by 100 trials of extinction

Responding to an extinguished CS has been shown to recover when that CS is presented in a context different from that in which it was extinguished (e.g., Bouton & Peck, 1989; Bouton & Ricker, 1994). AMAN can anticipate this context specificity of extinction via at least two mechanisms. One of these allows elements that form part of the extinction context to acquire some inhibitory strength that contributes to the loss (extinction) of responding to the CS in that context; this inhibition is then lost when the CS is presented in a different context after extinction, leading to some recovery of responding. However, this mechanism does not explain context-specific extinction under circumstances in which two CSs are extinguished in different contexts before the contexts are switched for test (Harris, Jones, Bailey, & Westbrook, 2000). We will return to this issue later.

In a feature-negative discrimination, $A+AX-$, the feature X acquires inhibitory strength that can reduce responding even when X is presented with a different CS, B (Pavlov, 1927). The model explains the acquisition of conditioned inhibition as the increase in negative associative strength between the E units of X and the US. However, unlike the Rescorla–Wagner (1972) model, AMAN does not predict extinction of inhibition when an inhibitor is presented alone, without the US. This is because the E units of the US cannot become negatively activated, and therefore there is no rise in E units of the US when an inhibitor is presented alone, and thus there is no opportunity for any increment in strength between the E units of the inhibitor and the US.

Summation As described by Harris and Livesey (2010), AMAN predicts the summation of conditioned responses when two separately conditioned CSs are presented simultaneously as a compound (Andrew & Harris, 2011; Kehoe, 1982; Rescorla, 1997). This is because, like many associative models, the conditional activation of the US representation is assumed to be based on summed associative input from each CS present on that trial. However, the amount of summation between two CSs will be reduced to the extent that they suppressively normalize one another's activation. Since normalization is assumed to be greater between stimuli that belong to the same modality than for stimuli from different modalities, this accounts for the observation that summation is greater between CSs from different modalities than between CSs from a single modality (Kehoe, Horne, Horne, & Macrae, 1994; Thein, Westbrook, & Harris, 2008).

Cue competition The summing of associative strengths affects both responding and learning, and therefore the model explains a variety of effects that demonstrate the competition of cues during learning. For example, AMAN is equivalent to other models that use a summed error term (e.g., Rescorla & Wagner, 1972) as the means by which they explain the following keystone phenomena: overshadowing (weaker conditioning to a CS when it is conditioned in compound with another stimulus than when it is conditioned in isolation); blocking (weaker conditioning to a CS when it is conditioned in compound with a previously conditioned CS than when it is conditioned in compound with a neutral stimulus); unblocking by increasing the magnitude of the US (Kamin, 1968); the US preexposure effect (weaker conditioning when a US has been repeatedly presented on its own, prior to CS–US pairings; Randich & LoLordo, 1979; Wagner, 1969); and relative validity (superior conditioning to a CS when it is conditioned in compound with a stimulus that is less well correlated with the US; Wagner, Logan, Haberlandt, & Price, 1968). As such, it should not be surprising that the model accounts for both the decrease in responding when two separately conditioned CSs are conditioned in compound (“overexpectation”) and the facilitation of learning about a CS when it is paired with a conditioned inhibitor during conditioning (“superconditioning”) (Garfield & McNally, 2009; Lattal, 1998; Pearce & Redhead, 1995; Rescorla, 1971). The model can also account for faster learning with longer intertrial intervals (Gibbon, Baldock, Locurto, Gold, & Terrace, 1977; Spence & Norris, 1950) in terms of differential extinction of the context over the intertrial interval, thereby permitting faster discrimination between the presence and absence of the CS.

Generalization Like other elemental models, the Harris and Livesey (2010) model predicts that conditioned responding to one CS will generalize to a second stimulus to the extent

that the two stimuli share common elements (Guttman & Kalish, 1956). It also uses a learning rule that functions in a fashion similar to a summed error term, and thus provides a means to achieve perfect performance on a simple discrimination (S + vs. S–), despite common elements shared between the two stimuli (Pavlov, 1927), because elements unique to S– acquire inhibitory strength that cancels the generalized excitatory strength from S+. (These mechanisms have also been shown to account for peak shift effects following discrimination training of S + vs. S–; Blough, 1975; Hanson, 1959; Terrace, 1968.)

Sensory preconditioning and second-order conditioning Finally, like many similar elemental models, AMAN explicitly assumes that excitatory and inhibitory associations are acquired between CSs, as well as between a CS and a US. Therefore, the model is capable of accounting for evidence of sensory preconditioning (responding to a stimulus that was initially paired with another stimulus before the latter was conditioned) and second-order conditioning (responding to a stimulus that was paired with a CS after the latter had been conditioned; Brogden, 1939; Rizley & Rescorla, 1972). Nonetheless, in either case, the activation of US elements by a second-order CS is weak, because it depends on associative activation of E units of either the primary CS (Rizley and Rescorla showed that the activation of the primary CS contributes little to the responding to the second-order CS) or the US, and such associative activation is much weaker than direct activation by the stimulus itself.

Phenomena arising from the interactions between elements

AMAN describes mechanisms by which stimulus elements interact to influence one another's activation. As described earlier, elements suppress (normalize) the activity of E units of other elements both by direct activation of inhibitory units and via the normalizing influences between their attention units. E units within and between CSs also develop excitatory and inhibitory associations as a consequence of stimulus exposure. As reviewed below, these influences allow stimuli to interact in ways that enable the model to explain a number of empirical findings that are otherwise beyond the scope of simple elemental models.

External inhibition and overshadowing Suppressive normalization between elements provides a mechanism for external inhibition (the reduction in responding to a CS when it is paired with a neutral stimulus; Pavlov, 1927). Elements of the added stimulus will reduce activation of the US elements by CS elements, and therefore reduce conditioned responding, by two means: (1) by directly increasing suppression of the associatively activated US elements and

(2) by indirectly reducing associative activation of the US elements by increasing suppression of the CS elements. The suppressive normalization of CS and US elements is also responsible for one-trial overshadowing: The activation strength of each CS and US element is reduced when a compound of two CSs is presented rather than a single CS (James & Wagner, 1980; Mackintosh & Reese, 1979). However, the loss of conditioning that results from this competitive normalization is of smaller magnitude than the competitive interactions that take place between CSs when they vie for associative strength during conditioning of a compound in a standard overshadowing procedure (as determined by the competitive-learning rule, equivalent to that defined by Rescorla & Wagner, 1972). This is in line with empirical evidence related to the comparison between adding and subtracting a stimulus (see Brandon, Vogel, & Wagner, 2000, for empirical evidence related to the comparison between adding vs. subtracting a stimulus). Examples of these differences are shown in Fig. 3.

Nonlinear discriminations The normalization between CS elements was originally invoked to allow the simple elemental approach to solve nonlinear discriminations in which the associative strength of the CS compound does not equal

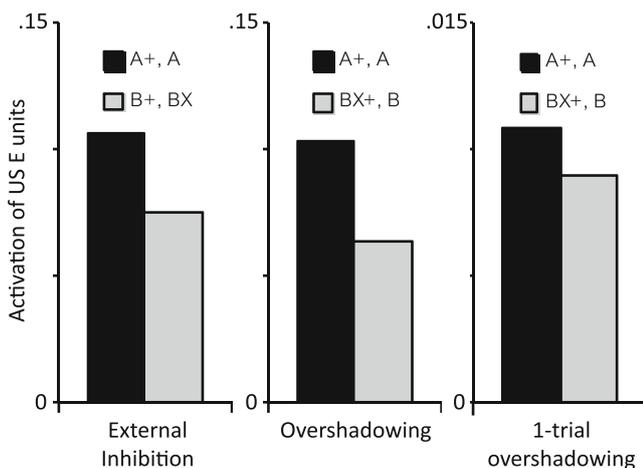


Fig. 3 Left: Simulation of external inhibition when a stimulus (X) is added to a CS (B), after B has been conditioned on its own (B+). The activation of US elements by BX is weaker than that by a comparison CS, A, that had been conditioned in the same manner as B but is tested without X. Center: Simulation of overshadowing between two stimuli, B and X, that have been conditioned together as a compound (BX+) before B is tested on its own. On test, B activates US elements more weakly than does a comparison CS, A, that was conditioned alone. Right: Simulation of one-trial overshadowing when B is tested alone after a single BX+ conditioning trial, compared with performance to A after a single A+ conditioning trial. Note that the scale on the y-axis is an order of magnitude smaller on this right plot than on either of the other plots (which are equal), because simulated conditioning strength is much weaker after just one conditioning trial when all other parameters are kept equal across simulations

the summed associative strength of the individual CSs, such as negative patterning (A+, B+, AB−) and the biconditional discrimination (AB+, CD+, AC−, BD−). The way that this is achieved, and its success in dealing with the more complex patterning discriminations tested by Pearce and colleagues (e.g., Pearce, 1994), has been dealt with at length by Harris (2010) and Harris and Livesey (2010) and will not be discussed in more detail here. Nonetheless, it is instructive to consider two empirical findings on this topic that the model does not account for as successfully as its forerunner (Harris, 2006). The current implementation of the Harris and Livesey (2010) model does not predict that the biconditional discrimination would be more difficult than negative patterning, as shown by Harris and Livesey (2008) and Harris, Livesey, Gharaei, and Westbrook (2008), and it incorrectly predicts that a redundant cue will facilitate rather than impede the acquisition of negative patterning (Pearce & Redhead, 1993; Rescorla, 1972).

Unequal learning to two CSs As described by Harris (2006), one important feature of a simple elemental model like AMAN is that it can explain some recent reports by Rescorla showing unequal learning to each of two CSs conditioned in compound if those CSs start with different conditioning strengths (Rescorla, 2000, 2001, 2002). For example, in a blocking design in which stimulus A is conditioned prior to conditioning of the compound AB, Rescorla (2001) showed that less is learned about A than about B during those compound training trials. This finding is problematic for associative models that assume that any learning that takes place during the AB conditioning trials will be equally shared by A and B. The elemental model initially proposed by Harris (2006), and subsequently developed by Harris and Livesey (2010), can account for Rescorla's (2001) result. To explain how this works, consider a very simplified example in which A, B, and the US each comprise three elements, and in which each element of A and B is connected to just one element of the US (this is a simplified illustration of the partial connectivity assumed to exist between CS and US elements). During the initial conditioning, each element of A acquires associative strength with one element of the US. When A is subsequently presented in compound with B, one of its elements is suppressed by B (and one, but not necessarily the same one, of B's elements is suppressed by A). Thus, some of A's associative strength is lost (as described already to account for external inhibition), and one US element is now available for further conditioning. The important point is that any conditioning that can occur during AB trials is confined to that one US element. However, because this US element is connected to the one A element that has been suppressed, A has no opportunity to increase its associative strength with the US. In contrast to this situation for A, there is no

preordained match between B's suppressed element and the relevant US element, and therefore B is in a better position than A to acquire whatever associative strength is available. Our simulations with AMAN confirm that this outcome holds even when each stimulus is represented with a larger number of elements and the connections between the E units of the elements are random rather than one to one.

Context-specific extinction We have already noted that the AMAN model can account for the context specificity of extinction by allowing elements of the extinction context to acquire inhibition of the US. The model can, under certain circumstances, also account for context-specific extinction in which two CSs, A and B, are extinguished in different contexts (Y and Z, respectively), before both are tested in Context Y (Harris et al., 2000). The evidence that responding to B recovers when tested in Y (rather than its extinction context, Z) cannot be explained by differential inhibitory learning about the contexts, because both contexts would have had the same opportunity to develop inhibition toward the US. AMAN can simulate greater responding to B than to A when tested in Context Y (see Fig. 4) by allowing inhibition to develop between each context and the CS extinguished in that context (i.e., between Y and A, and between Z and B), such that B is more active, and thus better able to activate elements of the US, in the other context (Y). However, our simulations have only succeeded in producing sufficient inhibition to achieve this result when each context has contained a large number of

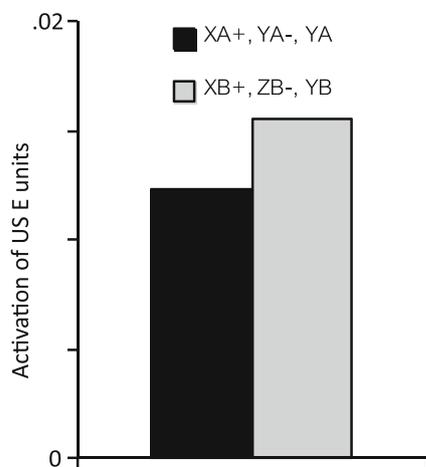


Fig. 4 Simulation of the context specificity of extinction, as demonstrated by Harris et al. (2000). In this simulation, two CSs, A and B, received simulated conditioning (+) in Context X, before each CS was extinguished (–) in a separate context (A was extinguished in Context Y, B was extinguished in Context Z). Finally, both CSs were tested in Context Y. On this test, B activated the US elements more strongly than did A because inhibitory associations between Y and A reduced activation of A's own elements, which in turn reduced the associative activation of the US

elements relative to the number defining each CS. For example, in the simulations shown in Fig. 4, each context contained three times as many elements as each CS. We believe that this requirement is justifiable, given the greater spatial and sensory complexity of contextual stimuli, which is often deliberately enhanced in procedures that require several contexts to be discriminated from one another.

Occasion setting It was noted earlier that AMAN can account for conditioned inhibition in a simultaneous feature-negative discrimination, A+ versus AX–, in much the same way that the Rescorla–Wagner (1972) model does. By describing the activation of elements in “real time,” AMAN is also able to explain the occasion-setting properties of a stimulus in a serial feature-negative or feature-positive discrimination when X is presented in advance of A. An important property of the way that animals solve these serial discriminations is that the feature X does not acquire an excitatory or inhibitory association with the US, but controls the expression of A's excitatory or inhibitory associative strength (Holland, 1984; Ross & Holland, 1981). Indeed, Holland (1991) showed that a single stimulus, X, can concurrently serve as an excitatory occasion setter ($X \rightarrow A +$ vs. $A -$) and as an inhibitory occasion setter ($X \rightarrow B -$ vs. $B +$). AMAN can achieve this result if X alters the activation of A's elements on $X \rightarrow A$ trials, and alters B's elements on $X \rightarrow B$ trials, in such a way that “A after X” is discriminable from “A after nothing” and, likewise, “B after X” is discriminable from “B after nothing.” X can exert this influence over A's and B's elements both through direct suppressive normalization and by the acquisition of associative connections between X's elements and some of A's and B's elements. Because X changes the activity of A's and B's elements, some elements of each CS can acquire excitatory associative strength, while other elements of that CS can acquire inhibitory strength. The output of one such simulation is presented in Fig. 5. Needless to say, the effect relies on X being close enough in time to A and B that X's elements are still sufficiently active to influence A's and B's elements. In order to simulate the same serial occasion-setting result, but across a long trace interval separating X from A and B, one would need to select slower decay parameters than we have adopted in the simulations presented throughout this article.

Timing of the conditioned response One noteworthy property of the Harris and Livesey (2010) model that has not been previously explored is that it can simulate one aspect of the temporal specificity of conditioned responding when conditioning involves a fixed CS–US interval. Many demonstrations have shown that, when animals are trained with a fixed CS–US interval, the strength of the conditioned response rises as time elapses during the CS, to peak at the approximate time of the US (Davis, Schlesinger, &

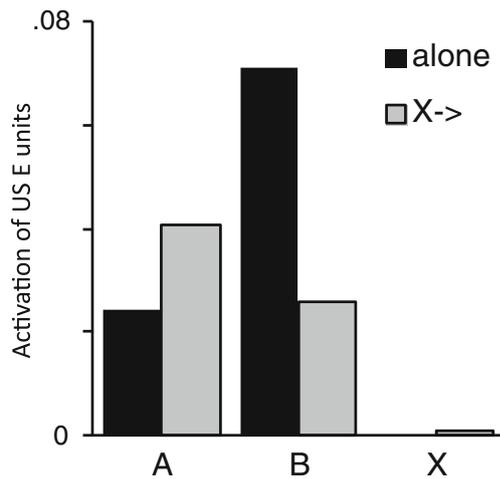


Fig. 5 Simulation of conditioned response strength during two CSs, A and B, when each was presented alone or presented immediately after a serial occasion-setting stimulus, X (after Holland, 1991). In this simulation, the model was trained with two concurrent discriminations: A vs. X→A+, and B+ vs. X→B-. All three stimuli had input strengths between 0.5 and 1, and inputs lasted for five time units (Time Points 1 to 5 for X, 6 to 10 for A and B), and the intertrial interval was 38 time units. The plot shows the average activation strength of the US E units during the presentations of A and B after 500 cycles of training, and also shows US activation during presentation of X. The model successfully produced greater US activation during A when it followed X than when it was presented alone, and greater US activation during B when it was presented alone than when it followed X

Sorenson, 1989; Kehoe & Joscelyne, 2005; Pavlov, 1927; Roberts, 1981; Smith, 1968; Williams, Lawson, Cook, Mather, & Johns, 2008). Moreover, this timing of the conditioned response typically shows scalar invariance, in that variability in the response timing around the moment of US presentation is a fixed proportion of the length of the CS–US interval (Gibbon, 1977). As illustrated in Fig. 6, some of these properties emerge in simulations of AMAN. When the model is “trained” with different fixed CS–US intervals, US activation by the CS increases as time elapses during the CS, and the rate of this rise is scaled to the length of the CS–US interval. This occurs as a result of differences in the distributions of excitatory and inhibitory strength among different CS elements. E units of some CS elements, which are active at the onset of the CS but lose activation as time elapses during the CS, become inhibitors of the US, because these elements are only active when the US is associatively activated and become inactive when the US occurs. Excitatory associative strength is only acquired by the CS elements that remain active at the time of the US presentation.

The mechanism that we have just described to explain the timing of conditioned responses is clearly very similar to the inhibition of delay described by Pavlov (1927) and formalized in more recent time-derivative elemental models (e.g., Ludwig, Sutton, & Kehoe, 2008; Sutton & Barto, 1981, 1990). However, it is important to recognize that the

differences in the time courses of element activations that allow for timing of conditioned responses have not been explicitly programmed into AMAN. Instead, different elements develop different temporal profiles by virtue of the changing landscape of suppressive normalization between

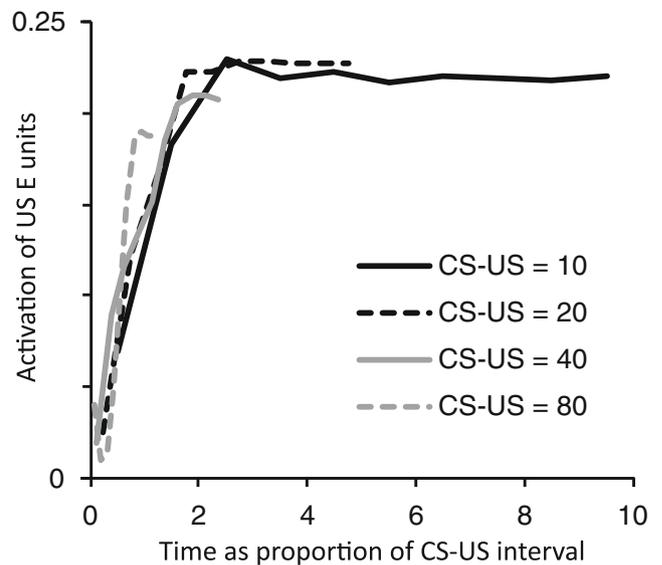
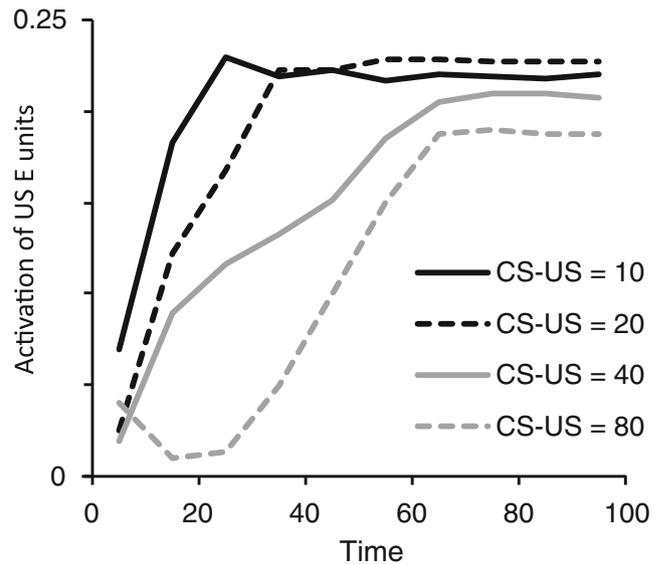


Fig. 6 The expected strength of conditioned responding (given by the activation of US E units) as time elapses during each of four CSs. During simulated conditioning, the CSs had fixed durations (10, 20, 40, or 80 time steps) and terminated with the presentation of the US. The top plot shows conditioning strength after 500 conditioning trials, during a test in which each CS was presented for 100 time steps. It is clear that, for each CS, US activation increases as time elapses, reaching a peak soon after the expected time of the US (corresponding to the CS duration during conditioning). The slope of this rise in US activity varies in proportion to the CS–US interval. This scalar property is revealed in the lower plot, which shows that the simulated data for each CS are superimposed when time is rescaled as a proportion of the CS–US interval

elements, as well as the development of excitatory and inhibitory associations between elements. At CS onset, activity in all elements rises in response to sensory input, but as time elapses, the E units become suppressed as a part of the simultaneously rising normalization among the elements, and the weakest E units are the first to decay. With repeated experience, this process becomes more pronounced. The weaker elements become associatively inhibited by the stronger elements, in addition to their direct suppression as part of normalization. The stronger elements, on the other hand, increase their activation strength as time elapses, due to excitatory associative connections among their E units. This change in the time course of activation of CS E units is shown in Fig. 7, after a small number of trials (left plot), and again after many trials (right plot).

Two limitations must be acknowledged in regard to the way that AMAN deals with the timing of conditioned responses. First, the simulations presented here have all

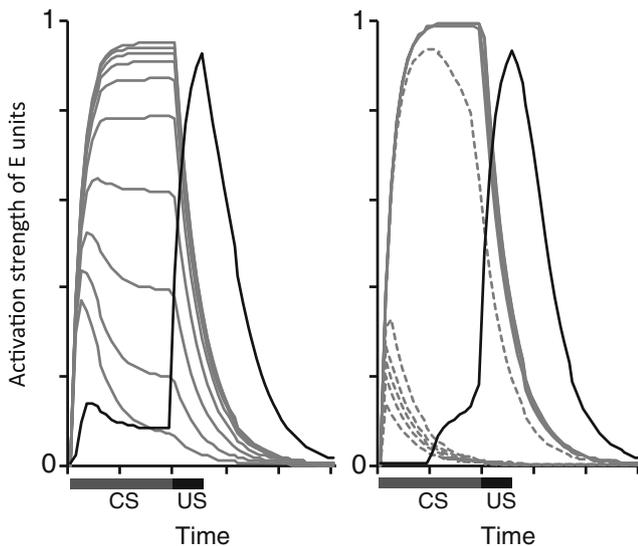


Fig. 7 Activation of ten CS E units (gray lines; sensory input increasing uniformly from 0.5 to 1) and one representative US E unit (black line) as time elapses during a trial in which the CS is present for the first 20 time steps and the US is present for the next five time steps. The left plots shows these activation patterns after relatively few (10) simulated CS–US trials; the right plot shows the activation patterns after many (750) trials. After only a few training trials, all CS E units are activated when the US is presented, and therefore they develop positive associative connections with the E units of the US, but the strength of these connections scales with their activation strength (the most active CS elements have the strongest associative strength). The result is that the CS activates the US elements at a more-or-less uniform level across the duration of the CS. With continued conditioning, the weaker elements of the CS become suppressed early on. This promotes extinction of their connection and finally converts them into a conditioned inhibitor (inhibitory elements are identified here by dashed lines in the right plot). As a result, they inhibit activation of the US during the initial period of the CS, such that the US elements are only activated shortly before the expected time of the US itself

involved delay conditioning in which the US is presented when the CS terminates. The model does not perform well in simulating conditioning when the US is presented early during a prolonged CS. Such schedules produce a large backward pairing (CS after US) that undermines the development of net excitatory associative strength between the CS and the US in the model's simulations. Second, while AMAN can effectively simulate inhibition of delay (Pavlov, 1927), we note that it is not able to simulate the temporal specificity of the conditioned response beyond the time of the US (Davis et al., 1989; Kehoe & Joscelyne, 2005; Roberts, 1981; Smith, 1968; Williams et al., 2008). For example, if rats are conditioned with a US that occurs midway through a long CS, the conditioned response not only rises during the CS as the scheduled time of the US approaches, but it also falls as time passes beyond the scheduled time of the US, as assayed on probe trials in which the US is omitted (e.g., Williams et al., 2008). It is clear from Fig. 6 that this fall in the conditioned response beyond the scheduled time of the US is not reproduced by AMAN; instead, the CS continues to activate US elements well beyond the expected US time. Our analysis suggests that this is because, as time elapses, the activation profile of CS elements becomes more stable, thereby reducing the opportunity for inhibitory strength to be acquired by CS elements that are more strongly activated after the US than during it.

Latent inhibition Finally, Harris and Livesey (2010) showed that AMAN can simulate the latent inhibition that results from repeated presentations of a stimulus prior to its conditioning, as well as superior discrimination between two CSs with overlapping elements if the CSs have been preexposed prior to training (so-called *perceptual learning*). However, the model does not account for differences in perceptual learning when those CS preexposures are intermixed (A . . . B . . . A . . . B . . .) rather than blocked (A . . . A . . . A . . . B . . . B . . . B).

Challenges

A number of phenomena are not explained by AMAN. Some of these describe aspects of conditioned responding that are simply outside the scope of any of the constructs used by the model to capture learning. For example, because the model identifies CS–US conditioning with the associative activation of US elements, rather than with changes in the CS elements, it does not explain how different CSs might influence the form of conditioned responses (Holland, 1977), nor does it account for changes in the associability of CSs in relation to their history of pairing with the US (e.g., Hall & Pearce, 1979). And, like the associative models that it borrows from, its commitment to predicting overshadowing prevents AMAN from being able to explain potentiation

(the enhancement of conditioning to a CS when it is reinforced in compound; Batson & Batsell, 2000; Westbrook, Clarke, & Provost, 1980; Westbrook, Homewood, Horn, & Clarke, 1983), and its commitment to the mechanisms that produce blocking prevents it from explaining evidence for unblocking by a decrease in US magnitude (Dickinson & Mackintosh, 1979; Dickinson, Mackintosh, & Cotton, 1980). In much the same vein, because its learning mechanisms are tuned to positive or negative contingencies between CS and US, it does not explain what learning takes place when the CS and US are truly uncorrelated (Bonardi & Hall, 1996). Finally, the model is not equipped with any process that can account for the effect of the passage of time in recovering conditioned responding after extinction (Pavlov, 1927) or following latent inhibition (Westbrook, Jones, Bailey, & Harris, 2000).

These challenges notwithstanding, we have shown here that the conditioning model proposed by Harris and Livesey (2010) is well equipped to explain a large number of empirical observations regarding Pavlovian conditioning and discrimination learning. In this sense, the model clears many hurdles. But developing a model of this type does require specifying many details that may make the model appear overly elaborate, or even cumbersome. Our intention here has been to take simple concepts that underlie an elemental approach to conditioning and to develop these to a natural conclusion, and the present exercise shows how such a model fares in the broad experimental arena. We believe that the model has acquitted itself well in this showjumping exercise, even if it were not to score full marks in the dressage.

Appendix: The parameter definition file used in all simulations described in this article

Note: Lines shown here in italics are not read by the program files and are included below purely for explanatory purposes. In the parameter file uploaded into the simulator, these lines are marked by a # symbol at the head of each line. The parameter definitions are shown in bold.

1. PARAMETERS AND OPTIONS

Each of the following lines consists of the name of a parameter/option, an equal sign, and a number—all separated by a space.

Please do not change the names of the parameters/options.

The notations included in brackets within the comments sections refer to those used in Harris and Livesey (2010).

1.1 BASIC STRUCTURE OF THE NETWORK AND THE SIMULATION

Number of time steps in each trial. Please make sure that you specify enough time steps to present all stimuli and record all activation as defined in the trial definition. Also bear in mind that stimuli decay gradually, i.e. a stimulus' units may still be active after its offset.

We advise allowing at least 25 ticks beyond the offset of the last stimulus.

num_TicksPerTrial=50

Number of elements representing each stimulus. Each stimulus specified in the stimulus definition is represented by several elements that in turn consist of one E, I, and A unit each. The activation of each of these units and their associations with each other have to be updated in each time step using several iterations. Increasing the number of elements increases computing time substantially.

num_elements=10

Connectivity between E units: This parameter specifies the likelihood that one E unit is connected to another E unit. The likelihoods of the two unidirectional connections between two E units are independent. Increasing the likelihood increases the number of connections between E units, and thus the learning speed. A likelihood of .5 maximizes the variability across the network that is necessary to solve certain learning tasks.

If p is <1 , the output of any single simulation will vary, and it is therefore advisable to average across repeated simulations. For many simulated experiments, the connectivity does not have important consequences for the simulation, and thus we advise setting $p=1$ to avoid variability between simulations.

connect_p=1

1.2 PARAMETERS FOR ACTIVATION FUNCTION

Rule to calculate external sensory input (S_x) by a stimulus to its elements

Each stimulus activates a set of elements (or more precisely, its E and A units). Their activation is determined by the minimal and maximal activation of this stimulus as specified in the stimulus definition and the rule selected here.

The simulator allows the user to choose between three different rules to determine this external sensory input for each element:

1=linearly: element input strengths are graded in equal steps from maximal to minimal activation

2=randomly within range: draws for each element a value from the uniform distribution on the activation interval [minimum, maximum]

3 = constant: all elements are activated to the mean of the minimal and maximal activation

If option 2 is selected, the output of any single simulation will vary, and it is therefore advisable to average across repeated simulations. For many simulated experiments, there is little difference between options 1 and 2 for the simulation. In such cases, we advise selecting option 1 to avoid variability between simulations.

S_distribution=1

Parameter for response potential functions of E, I, and A units (Rpot):

Exponent (p) of response potential function (Rpot): Higher values of p make the response potential functions sigmoid and increase the maximal slope.

Power_E=2

Power_I=2

Power_A=5

Constant (D) of response potential function (Rpot):

This parameter affects the relative sensitivity of the response potential to an input and prevents the denominator ever being zero

D=0.04

Rate parameters (delta) for change of the activation function (dR/dt): These two parameters control how fast the activations of E, I, and A units actually rise and decay, respectively.

Rate parameter (delta) for rising of the activation (R), i.e. when the unit's response potential to its current input is higher than its current activation [(Rpot - R) > 0]

Growth_E=0.5

Growth_I=0.5

Growth_A=0.5

Rate parameter (delta) for decay of the activation (R), i.e. when the unit's response potential to its current input is lower than its current activation [(Rpot - R) < 0]

Decay_E=0.2

Decay_I=0.2

Decay_A=0.2

1.3 FIXED CONNECTION WEIGHTS

1.3.1 Normalization based on similarity

Connection weight between E and I units of different elements (z): E units activate other elements' I units, and thus increase the inhibition of E units by I units, i.e. normalization in the network. Instead of a graded change in values of z across elements, three different degrees of normalization can be specified:

1. Normalization of I units by E units of elements belonging to other stimuli.

norm_across_all=0.01

2. Normalization of I units by E units of elements of other stimuli that share the same modality. The modality is

specified in the stimulus definition file. Stimuli with a modality of 0 do not share a modality with any other stimulus. This value will be divided by the number of elements in each stimulus.

norm_within_modality=0.5

3. Normalization of I units by E units of other elements of the same stimulus. This value will be divided by the number of elements in the stimulus.

norm_within_stim=0.5

1.3.2 Normalization based on attentional field:

Connection weight from A to I unit (ka): A units inhibit I units of the same elements. Increasing the connection weight therefore decreases normalization of the E units. The parameter value should be positive (even though the connection is inhibitory).

Weight_IA=4

Connection weight between A units (w): A units are in turn inhibited by other elements' A units. Increasing this connection weight therefore increases normalization of E units. The parameter value should be positive (even though the connection is inhibitory).

Weight_AA=0.1

1.4 LEARNING PARAMETER

1.4.1 Parameter for coactivation function of two elements (DELTA):

Changes in the connection weights between E units (i.e. the associative strength V) depend on the coactivation of the two elements. The coactivation changes in response to increments in the recipient's E and I and is scaled by two rate parameters, betaE and betaI. As the default, both parameters are set to 0 when the activations of E and I are decaying, so that associative changes are driven only by increases in the elements' activation. Increasing the parameter for rising activation accelerates learning.

Parameter (betaE) for the effects of rising of recipient's E unit's activation (dE_dt > 0)

Beta_E_pos=0.02

Parameter (betaE) for the effects of decay of recipient's E unit's activation (dE_dt < 0)

Beta_E_neg=0.0

Parameter (betaI) for the effects of rising of recipient's I unit's activation (dI_dt > 0)

Beta_I_pos=0.1

Parameter (betaI) for the effects of decay of recipient's I unit's activation (dI_dt < 0)

Beta_I_neg=0.0

1.4.2 Rate parameter (kV) for changes in the connection weight function (dV_dt)

This parameter controls how fast associative weights are actually changing. The parameter value should be between

0 and 1. High values accelerate learning; low values result in more stable learning.

k_V=1

Rate parameter (k_{α}) for changes in associability of sender's E unit function ($d_{\alpha dt}$)

This parameter controls how fast associability is actually changing. The parameter should have values between 0 and 1.

k_Alpha=1

1.5 OUTPUT OPTIONS

1.5.1 Figures:

Which figures should be created? 0=no; 1=yes:

Final associative weights (i.e. connections between E units) at the end of the experiment

fig_V=0

Activation of all E units in each time step

fig_E=0

Activation of all I units in each time step

fig_I=0

Activation of all A units in each time step

fig_A=0

E units' activation, recorded and averaged for each trial, as specified in the trial definition

fig_CR=1

1.5.2 Output: how should the E units' activation, recorded and averaged for each trial as specified in the trial definition, be saved?

0=not at all,

1=as xls: each phase is saved in a separate sheet with the corresponding name

2=as a MATLAB mat-file: each phase is saved in a cell of a cell array

mat_CR=0

References

- Andrew, B. J., & Harris, J. A. (2011). Summation of reinforcement rates when conditioned stimuli are presented in compound. *Journal of Experimental Psychology: Animal Behavior Processes*, 37, 385–393.
- Batson, J. D., & Batsell, W. R., Jr. (2000). Augmentation, not blocking, in an A+/AX+flavor-conditioning. *Psychonomic Bulletin & Review*, 7, 466–471.
- Blough, D. S. (1975). Steady state data and a quantitative model of operant generalization and discrimination. *Journal of Experimental Psychology: Animal Behavior Processes*, 104, 3–21.
- Bonardi, C., & Hall, G. (1996). Learned irrelevance: No more than the sum of CS and US preexposure effects? *Journal of Experimental Psychology: Animal Behavior Processes*, 22, 183–191. doi:10.1037/0097-7403.22.2.183
- Bouton, M. E., & Peck, C. A. (1989). Context effects on conditioning, extinction, and reinstatement in an appetitive conditioning preparation. *Animal Learning & Behavior*, 17, 188–198.
- Bouton, M. E., & Ricker, S. T. (1994). Renewal of extinguished responding in a second context. *Animal Learning & Behavior*, 22, 317–324.
- Brandon, S. E., Vogel, E. H., & Wagner, A. R. (2000). A componential view of configural cues in generalization and discrimination in Pavlovian conditioning. *Behavioural Brain Research*, 110, 67–72.
- Brogden, W. J. (1939). Sensory pre-conditioning. *Journal of Experimental Psychology*, 25, 223–232.
- Davis, M., Schlesinger, L. S., & Sorenson, C. A. (1989). Temporal specificity of fear conditioning: Effects of different conditioned stimulus-unconditioned stimulus intervals on the fear-potentiated startle effect. *Journal of Experimental Psychology: Animal Behavior Processes*, 15, 295–310.
- Dickinson, A., & Mackintosh, N. J. (1979). Reinforcer specificity in the enhancement of conditioning by posttrial surprise. *Journal of Experimental Psychology: Animal Behavior Processes*, 5, 162–177.
- Dickinson, A., Mackintosh, N. J., & Cotton, M. M. (1980). Surprise and blocking: Effects of the number of compound trials. *Animal Learning & Behavior*, 8, 387–391.
- Garfield, J. B. B., & McNally, G. P. (2009). The effects of FG7142 on overexpectation of Pavlovian fear conditioning. *Behavioral Neuroscience*, 123, 75–85.
- Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*, 84, 279–325.
- Gibbon, J., Baldock, M. D., Locurto, C., Gold, L., & Terrace, H. S. (1977). Trial and intertrial durations in autoshaping. *Journal of Experimental Psychology: Animal Behavior Processes*, 3, 264–284.
- Grossberg, S. (1973). Contour enhancement, short-term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics*, 52, 213–257.
- Guttman, N., & Kalish, H. I. (1956). Discriminability and stimulus generalization. *Journal of Experimental Psychology*, 51, 79–88.
- Hall, G., & Pearce, J. M. (1979). Latent inhibition of a CS during CS-US pairings. *Journal of Experimental Psychology: Animal Behavior Processes*, 5, 31–42. doi:10.1037/0097-7403.5.1.31
- Hanson, H. M. (1959). Effects of discrimination training on stimulus generalization. *Journal of Experimental Psychology*, 58, 321–334.
- Harris, J. A. (2006). Elemental representations of stimuli in associative learning. *Psychological Review*, 113, 584–605.
- Harris, J. A. (2010). The arguments of associations. In N. A. Schmajuk (Ed.), *Computational models of conditioning* (pp. 53–70). Cambridge, U.K.: Cambridge University Press.
- Harris, J. A., & Carpenter, J. S. (2011). Response rate and reinforcement rate in Pavlovian conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, 37, 375–384.
- Harris, J. A., Jones, M. L., Bailey, G. K., & Westbrook, R. F. (2000). Contextual control over conditioned responding in an extinction paradigm. *Journal of Experimental Psychology: Animal Behavior Processes*, 26, 174–185.
- Harris, J. A., & Livesey, E. J. (2008). Comparing patterning and biconditional discriminations in humans. *Journal of Experimental Psychology: Animal Behavior Processes*, 34, 144–154.
- Harris, J. A., & Livesey, E. J. (2010). An attention modulated associative network. *Learning & Behavior*, 38, 1–26.
- Harris, J. A., Livesey, E. J., Gharaei, S., & Westbrook, R. F. (2008). Negative patterning is easier than a biconditional discrimination. *Journal of Experimental Psychology: Animal Behavior Processes*, 34, 494–500.
- Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, 9, 181–197.
- Holland, P. C. (1977). Conditioned stimulus as a determinant of the form of the Pavlovian conditioned response. *Journal of Experimental Psychology: Animal Behavior Processes*, 3, 77–104.
- Holland, P. C. (1984). Differential effects of reinforcement of an inhibitory feature after serial and simultaneous feature negative discrimination training. *Journal of Experimental Psychology: Animal Behavior Processes*, 10, 461–475.

- Holland, P. C. (1991). Transfer of control in ambiguous discriminations. *Journal of Experimental Psychology: Animal Behavior Processes*, *17*, 231–248.
- James, J. H., & Wagner, A. R. (1980). One-trial overshadowing: Evidence of distributive processing. *Journal of Experimental Psychology: Animal Behavior Processes*, *6*, 188–205.
- Kamin, L. J. (1968). “Attention-like” processes in classical conditioning. In M. R. Jones (Ed.), *Miami symposium on the prediction of behavior: Aversive stimulation* (pp. 9–31). Miami, FL: Miami University Press.
- Kehoe, E. J. (1982). Overshadowing and summation in compound stimulus conditioning of the rabbit’s nictitating membrane response. *Journal of Experimental Psychology: Animal Behavior Processes*, *8*, 313–328.
- Kehoe, E. J., Home, A. J., Home, P. S., & Macrae, M. (1994). Summation and configuration between and within sensory modalities in classical conditioning of the rabbit. *Animal Learning & Behavior*, *22*, 19–26.
- Kehoe, E. J., & Joscelyne, A. (2005). Temporally specific extinction of conditioned responses in the rabbit (*Oryctolagus cuniculus*) nictitating membrane preparation. *Behavioral Neuroscience*, *119*, 1011–1022.
- Konorski, J. (1948). *Conditioned reflexes and neuron organization*. Cambridge, UK: Cambridge University Press.
- Lattal, K. M. (1998). Overexpectation in appetitive Pavlovian and instrumental conditioning. *Animal Learning and Behavior*, *26*, 351–360.
- Ludwig, E. A., Sutton, R. S., & Kehoe, E. J. (2008). Stimulus representation and the timing of reward-prediction errors in models of the dopamine system. *Neural Computation*, *20*, 3034–3054.
- Mackintosh, N. J., & Reese, B. (1979). One-trial overshadowing. *Quarterly Journal of Experimental Psychology*, *31*, 519–526.
- McLaren, I. P. L., & Dickinson, A. (1990). The conditioning connection. *Philosophical Transactions of the Royal Society: Series B*, *329*, 179–186.
- 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.
- Pavlov, I. P. (1927). *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex*. (G. V. Anrep, Trans.). New York, NY: Dover.
- Pearce, J. M. (1994). Similarity and discrimination: A selective review and a connectionist model. *Psychological Review*, *101*, 587–607. doi:10.1037/0033-295X.101.4.587
- 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. (1995). Supernormal conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, *21*, 155–165.
- Randich, A., & LoLordo, V. M. (1979). Associative and non-associative theories of the UCS preexposure phenomenon: Implications for Pavlovian conditioning. *Psychological Bulletin*, *86*, 523–548.
- Rescorla, R. A. (1971). Variation in the effectiveness of reinforcement and nonreinforcement following prior inhibitory conditioning. *Learning and Motivation*, *2*, 113–123.
- Rescorla, R. A. (1972). “Configural” conditioning in discrete-trial bar pressing. *Journal of Comparative & Physiological Psychology*, *79*, 307–317.
- Rescorla, R. A. (1997). Summation: Assessment of a configural theory. *Animal Learning & Behavior*, *25*, 200–209.
- Rescorla, R. A. (2000). Associative changes in exciters and inhibitors differ when they are conditioned in compound. *Journal of Experimental Psychology: Animal Behavior Processes*, *26*, 428–438.
- Rescorla, R. A. (2001). Unequal associative changes when exciters and neutral stimuli are conditioned in compound. *Quarterly Journal of Experimental Psychology*, *54B*, 53–68.
- Rescorla, R. A. (2002). Effect of following an excitatory–inhibitory compound with an intermediate reinforcer. *Journal of Experimental Psychology: Animal Behavior Processes*, *28*, 163–174.
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York, NY: Appleton-Century-Crofts.
- Reynolds, J. H., & Chelazzi, L. (2004). Attentional modulation of visual processing. *Annual Review of Neuroscience*, *27*, 611–647. doi:10.1146/annurev.neuro.26.041002.131039
- Reynolds, J. H., & Heeger, D. J. (2009). The normalization model of attention. *Neuron*, *61*, 168–185.
- Rizley, R. C., & Rescorla, R. A. (1972). Associations in second-order conditioning and sensory preconditioning. *Journal of Comparative & Physiological Psychology*, *81*, 1–11.
- Roberts, S. (1981). Isolation of an internal clock. *Journal of Experimental Psychology: Animal Behavior Processes*, *7*, 242–268.
- Ross, R. T., & Holland, P. C. (1981). Conditioning of simultaneous and serial feature-positive discriminations. *Animal Learning & Behavior*, *9*, 293–303.
- Smith, M. C. (1968). CS–US interval and US intensity in classical conditioning of the rabbit’s nictitating membrane response. *Journal of Comparative and Physiological Psychology*, *66*, 679–687.
- Spence, K. W., & Norris, E. B. (1950). Eyelid conditioning as a function of the inter-trial interval. *Journal of Experimental Psychology*, *40*, 716–720.
- Sutton, R. S., & Barto, A. G. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review*, *88*, 135–171.
- Sutton, R. S., & Barto, A. G. (1990). Time-derivative models of Pavlovian reinforcement. In M. Gabriel & J. Moore (Eds.), *Learning and computational neuroscience: Foundations of adaptive networks* (pp. 497–537). Cambridge, MA: MIT Press, Bradford Books.
- Terrace, H. S. (1968). Discrimination learning, the peak shift, and behavioral contrast. *Journal of the Experimental Analysis of Behavior*, *11*, 727–741.
- 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.
- Wagner, A. R. (1969). Stimulus selection and a “modified continuity theory”. In G. H. Bower & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 3, pp. 1–41). New York, NY: Academic Press.
- Wagner, A. R. (2008). Evolution of an elemental theory of Pavlovian conditioning. *Learning & Behavior*, *36*, 253–265. doi:10.3758/LB.36.3.253
- Wagner, A. R., Logan, F. A., Haberlandt, K., & Price, T. (1968). Stimulus selection in animal discrimination learning. *Journal of Experimental Psychology*, *76*, 177–186.
- Westbrook, R. F., Clarke, J. C., & Provost, S. (1980). Long-delay learning in the pigeon: Flavor, color, and flavor-mediated color aversions. *Behavioral & Neural Biology*, *28*, 388–407.
- Westbrook, R. F., Homewood, J., Horn, K., & Clarke, J. C. (1983). Flavour–odour compound conditioning: Odour-potential and flavour-attenuation. *Quarterly Journal of Experimental Psychology*, *35B*, 13–33.
- Westbrook, R. F., Jones, M. L., Bailey, G. K., & Harris, J. A. (2000). Contextual control over conditioned responding in a latent inhibition paradigm. *Journal of Experimental Psychology: Animal Behavior Processes*, *26*, 157–173.
- Williams, D. A., Lawson, C., Cook, R., Mather, A. A., & Johns, K. W. (2008). Timed excitatory conditioning under zero and negative contingencies. *Journal of Experimental Psychology: Animal Behavior Processes*, *34*, 94–105.