

# An instance theory of associative learning

Randall K. Jamieson · Matthew J. C. Crump ·  
Samuel D. Hannah

Published online: 14 September 2011  
© Psychonomic Society, Inc. 2011

**Abstract** We present and test an instance model of associative learning. The model, Minerva-AL, treats associative learning as cued recall. Memory preserves the events of individual trials in separate traces. A probe presented to memory contacts all traces in parallel and retrieves a weighted sum of the traces, a structure called the *echo*. Learning of a cue–outcome relationship is measured by the cue’s ability to retrieve a target outcome. The theory predicts a number of associative learning phenomena, including acquisition, extinction, reacquisition, conditioned inhibition, external inhibition, latent inhibition, discrimination, generalization, blocking, overshadowing, overexpectation, superconditioning, recovery from blocking, recovery from overshadowing, recovery from overexpectation, backward blocking, backward conditioned inhibition, and second-order retrospective revaluation. We argue that associative learning is consistent with an instance-based approach to learning and memory.

**Keywords** Associative learning · Memory · Instance theory · Exemplar theory

---

R. K. Jamieson (✉)  
Department of Psychology, University of Manitoba,  
Winnipeg, MB, Canada R3T 2N2  
e-mail: randy\_jamieson@umanitoba.ca

M. J. C. Crump  
Brooklyn College - CUNY,  
Brooklyn, NY, USA

S. D. Hannah  
University of Queensland,  
Brisbane, Australia

In a simple associative learning procedure, a cue,  $A$ , is presented followed by an outcome,  $X$ . With experience, presentation of  $A$  elicits anticipation of  $X$ . The growth of anticipation is the process of associative learning.

Typically, associative learning is modeled as a gradual accrual of excitatory and inhibitory connections between stimulus units. Associative strength is treated as summative; so the strength of an association between stimuli on trial  $t$  stands for the entire history of learning. This scheme for understanding learning is well described in the Rescorla–Wagner model (Rescorla & Wagner, 1972) and the many theories that follow from it.

Whereas summative theories provide insight into learning, they make an unreasonable assumption. Namely, they deny that the learner remembers the events of separate learning trials—a problem that Miller, Barnet, and Grahame (1995) call the *assumption of path independence*. The problem is important. First, it distinguishes learning from memory. Second, it denies a growing body of evidence that animals other than humans remember the events of learning trials (e.g., Fagot & Cook, 2006; Vaughan & Greene, 1984; Voss, 2009).

In contrast to the classical theories of learning, instance theories of human memory identify the individual experience (i.e., the instance) as the primitive unit of knowledge and treat learning as the accumulation and deployment of instances from memory. Brooks (1978, 1987) was among the first to champion the view. Medin and Schaffer (1978) were among the first to formalize it (see also Reed, 1972). Hintzman’s (1986, 1988) Minerva 2 model and Nosofsky’s (1986) generalized context model are classic formalizations of the view. Kruschke’s (1992, 1996, 2001) and Logan’s (1988, 2002) models are modern expressions of the approach. Blough (1998) provided an instance-based formulation of associative learning based in signal detection theory.

In this article, we build on previous efforts (Jamieson, Hannah & Crump 2010) to show that an instance model of associative learning, Minerva-AL, can reproduce associative learning phenomena. We apply the model to a broad range of associative learning protocols to convince readers that memory principles merit consideration as explanatory devices of associative learning. To establish breadth, we tackle five points. First, we show that the model handles acquisition, extinction, and rapid reacquisition. Second, we show that Minerva-AL handles inhibition, including conditioned inhibition, latent inhibition, and external inhibition. Third, we show that the model handles cue competition, including blocking, overshadowing, superconditioning, and overexpectation. Fourth, we show that the model handles nonlinear discrimination problems, including negative patterning and biconditional discriminations. Fifth, we show that the model handles retrospective revaluation, including recovery from blocking, recovery from overshadowing, and second-order retrospective revaluation.

Each set of simulations demonstrates an aspect of the model. The demonstrations of acquisition, extinction, and rapid reacquisition give a clear picture of the model's mechanics. Demonstrations of conditioned inhibition show how the model comes to anticipate that an outcome will not occur—a corollary of discrepancy encoding. We will use the problem of cue competition to develop a clear depiction of cue interactions at encoding and retrieval—conditions that will be central to our solution to retrospective revaluation. We present simulations of discrimination to illustrate the model's ability to remember specific cue–outcome contingencies. We will use the problem of retrospective revaluation to illustrate complex interactions between encoding and retrieval mechanisms. The work shows that with a principled modification, a model of human memory can address the problem of animal learning.

### Minerva-AL

Minerva-AL follows from Hintzman's (1986, 1988) Minerva 2 model of human memory. Minerva 2 is an instance theory of human memory for explaining episodic recognition and judgment of frequency. The theory has since been applied to a wide range of memory phenomena (e.g., Arndt & Hirshman, 1998; Benjamin, 2010; Clark, 1997; Dougherty, Gettys, & Ogden, 1999; Goldinger, 1998; Hicks & Starns, 2006; Jamieson, Holmes & Mewhort 2010; Jamieson, Hannah & Crump 2010; Jamieson & Mewhort, 2009a, 2009b, 2010, 2011; Kwantes, 2005; Kwantes & Mewhort, 1999; Kwantes & Neal, 2006; Thomas, Dougherty, Sprenger, & Harbison, 2008).

Informally, Minerva 2 is a theoretic framework that describes the memorial processes involved in representing, storing, and retrieving instances of experience. The model assumes that each experience is stored in memory as a unique trace (i.e., an instance). Presenting a probe to memory retrieves an aggregate of traces similar to it. Because a probe retrieves traces similar to it, it retrieves a representation of itself from memory. This is how the model accomplishes recognition. Because a probe retrieves whole traces, and because traces include cues that co-occur, a probe that consists of a single cue also retrieves the events it has co-occurred with in the past. This is how the model accomplishes cued recall. We will use the model's mechanism for cued recall to model associative learning.

Formally, Minerva 2 is a computational theory of memory. In the model, a stimulus, or event, is represented by a vector of  $n$  elements (or features). These features can refer to either particular stimulus properties (e.g., has wings) or information states (e.g., neural potentials). The distinction is not important to the model predictions.

Each feature of a representation takes one of three discrete values: +1, -1, or 0. A value of +1 or -1 indicates that the feature is relevant to the stimulus description; a value of 0 indicates that the feature is indeterminate or irrelevant.

Co-occurrence of events is represented by summing event representations to form a single vector. For example, if two events  $A = [0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0]$  and  $B = [0, 0, 0, 0, 0, 0, 0, 1, 1, 0]$  co-occur, their co-occurrence is represented as  $AB = A + B = [0, 0, 0, 0, 0, 1, 1, 0, 0, 1, 1, 0]$ . Generally, representations are orthogonal:  $A$  shares no features with  $B$  (as in the example).

In Minerva 2, memory is a two-dimensional matrix,  $M$ . In the matrix, rows store instances or events, and columns represent the features in each event. An event,  $E$ , is encoded as a trace in memory by copying its vector to the next row in the matrix. The probability that each feature in the trace is stored to memory is a model parameter,  $L$ . Thus, as  $L$  increases from zero to one, an event is stored more completely in memory.

Probing memory with a cue activates all traces in proportion to their similarities to the cue. In Minerva 2, similarity of the probe,  $P$ , to trace  $i$  in memory,  $M_i$ , is computed as

$$S_i = \frac{\sum_{j=1}^n P_j \times M_{ij}}{n_R} \quad (1)$$

where  $P_j$  is the value of the  $j$ th feature in the probe,  $M_{ij}$  is the value of  $j$ th feature of the  $i$ th row in memory,  $n$  is the number of features in the vectors under comparison, and  $n_R$  is the number of nonzero features in the vectors under comparison. The measure behaves similarly to the Pearson

correlation coefficient. Similarity is +1 when the row is identical to the probe, -1 when the row is opposite to the probe, and 0 when the row is orthogonal to the probe.

Trace  $i$ 's activation,  $A_i$ , is a nonlinear function of its similarity to the probe,

$$A_i = S_i^3 \quad (2)$$

In principle, the probe activates all traces in memory. However, the activation function ensures that the traces that are most similar to the probe are activated most strongly.

The information that a probe retrieves from memory is another vector,  $C$ , called the *echo*. Each of the  $j$  elements in the echo is equal to the sum of the corresponding weighted elements in the  $i = 1 \dots m$  traces in memory,

$$C_j = \sum_{i=1}^m A_i \times M_{ij} \quad (3)$$

Hintzman (1986) illustrated how to use the echo to simulate cued recall. Let the vector elements  $j = 1 \dots k$  stand for a name and the elements  $j = k+1 \dots n$  stand for a face. To retrieve a face given a name, a probe is constructed that has features  $j = 1 \dots k$  filled in and features  $j = k+1 \dots n$  empty (i.e., filled with zeroes). Given that the name represented in features  $j = 1 \dots k$  finds a match in memory, features  $j = k+1 \dots n$  in the echo will approximate the features of the associated face (retrieval of a name given a face is accomplished in the opposite fashion).

Quality of cued recall—how well the retrieved vector approximates the target—is indexed in two steps: Values in the echo are normalized,

$$C'_j = \frac{C_j}{\max |C_{1,n}|} \quad (4)$$

then similarity is computed between the normalized echo and the relevant target associate,  $X$ :

$$X|P = \frac{\sum_{j=1}^n X_j \times C'_j}{n_R} \quad (5)$$

where  $P$  is the probe,  $X$  is the target associate, and  $n_R$  is the number of nonzero features in  $X$  and  $C'$ . The value  $X|P$  is read “retrieval of  $X$  given  $P$ .” The larger that  $X|P$  is, the better that  $X$  is retrieved by  $P$ . If the probe retrieves  $X$  perfectly, then  $X|P = 1$ . If the probe retrieves an inverse (i.e., opposing) representation of  $X$ , then  $X|P = -1$ . If the probe does not retrieve  $X$ , then  $X|P = 0$ .

Like most theories of human memory, Minerva 2 assumes independent encoding of items. For example, in a recognition memory experiment, each studied item is stored to a row in the memory matrix, without considering the order in which list items were presented. The same is true in studies of categorization and cued recall.

Whereas independent encoding of list items is sufficient to simulate performance in many memory experiments, it is insufficient to simulate learning. In a learning experiment, the problem of interest is to understand how memory for the events of previous trials influences encoding and memory on a present trial. Thus, to model learning, we adapted the Minerva 2 model so that the memory of a present trial is affected by memory of preceding trials. The adapted model is called Minerva-AL (AL for *associative learning*).

The principal difference between Minerva-AL and Minerva 2 is how the two models encode an experience. In the Minerva 2 model, memory for a trial is established by copying an event vector,  $E$ , to a row in memory. In Minerva-AL, memory is established for a trial by copying the difference between an event vector,  $E$ , and the echo,  $C$ , it retrieves. By encoding the difference between the event vector and the echo, memory of preceding trials (i.e.,  $C$ ) shapes encoding on a present trial. Critically, Minerva-AL's encoding operation emphasizes memory for unexpected over expected information in the event vector. For reasons that will become clear shortly, we will call the encoding operation *discrepancy encoding*.

Minerva-AL implements discrepancy encoding by subtraction. If  $M_{ij}$  is the  $j$ th feature of the  $i$ th trace in memory,  $E_j$  is the  $j$ th feature of the event vector, and  $C'_j$  is the  $j$ th feature of the normalized echo, then

$$M_{ij} = E_j - C'_j \quad (6)$$

with probability  $L$  ( $M_{ij} = 0$  with probability  $1 - L$ ).

To illustrate discrepancy encoding, consider a learning trial where  $A$  is presented followed by  $X$ . In our example,  $A$  and  $X$  are represented by four features, with  $A = [1.0, 1.0, 0.0, 0.0]$  and  $X = [0.0, 0.0, 1.0, 1.0]$ . Because the trial presents  $A$  followed by  $X$ , the event vector,  $E$ , is equal to  $E = A + X = [1.0, 1.0, 1.0, 1.0]$ . Suppose that on trial  $i$ ,  $A$  is presented and retrieves  $C' = [0.4, 0.1, 0.6, 1.0]$ . The values in the third and fourth elements of the echo (i.e., 0.6 and 1.0) show that the model retrieves a strong albeit imperfect expectation of  $X$ . According to discrepancy encoding (see Eq. 6), the information stored to row  $i$  in memory will equal  $M_i = E - C' = [1.0, 1.0, 1.0, 1.0] - [0.4, 0.1, 0.6, 1.0] = [0.6, 0.9, 0.4, 0.0]$ . Note that the most anticipated feature in  $C'$  (feature 4) is encoded as the smallest absolute value in  $M_i$  (i.e., 0.0), the second most anticipated feature in  $C'$  (feature 3) is encoded as the second smallest absolute value in  $M_i$  (i.e., 0.4), and so on. That is, Minerva-AL encodes the least anticipated features in  $E$  most strongly and the best anticipated feature not at all.

An important corollary of discrepancy encoding is that memory encodes *anticipated but unrepresented* events. To illustrate by extending the previous example, consider a next trial where  $A$  is presented without  $X$ . That is,  $E = A = [1.0, 1.0, 0.0, 0.0]$ . Because  $A$  was previously paired with  $X$ ,

$A$  retrieves  $C' = [0.4, 0.1, 0.6, 1.0]$ , indicating that  $A$  elicited an expectation of  $X$ . Even though  $X$  was not presented, the unfulfilled expectation for  $X$  is stored through discrepancy encoding as  $M_i = E - C' = [0.6, 0.9, -0.6, -1.0]$ . Note that, now, differently from the previous example, the information encoded to the third and fourth elements of  $M_i$  (i.e., the italicized features representing  $X$ ) take the opposite sign of the original representation for  $X = [0.0, 0.0, 1.0, 1.0]$ . This *inverse representation* of  $X$  records the fact that presenting  $A$  caused the model to expect  $X$  but that  $X$  did not occur. This property of the model is pivotal to our explanation of associative learning (Jamieson, Hannah & Crump 2010). Of course, we are not the first to argue for the importance of discrepancy encoding. Tolman (1932) discussed expectancy in learning. Kamin (1969) and von Restorff (1933) identified expectancy as a key principle of encoding and remembering. Whittlesea and Williams (2000, 2001a, 2001b) used violation of expectancy to explain memory-based inference. Rescorla and Wagner (1972) used surprise to model learning in cue competition. Finally, error correction algorithms in neural networks privilege unexpected over expected features (Gluck & Bower, 1988).

To implement discrepancy encoding, we augmented two computational aspects of the model. First, the similarity rule in Minerva 2 (see Eq. 1) is tailored to a situation where features of stimulus and memory representations can take one of only three discrete values  $\{+1, 0, -1\}$ . However, discrepancy encoding allows feature values to range continuously between  $\pm 2$ . The remedy was to replace Minerva 2's similarity formula with a standard vector cosine (see Kwantes, 2005, for precedence):

$$S_i = \frac{\sum_{j=1}^n P_j \times M_{ij}}{\sqrt{\sum_{j=1}^n P_j^2} \sqrt{\sum_{j=1}^n M_{ij}^2}} \quad (7)$$

where  $P_j$  is the value of the  $j$ th feature in the probe,  $M_{ij}$  is the value of the  $j$ th feature of the  $i$ th row in memory, and  $n$  is the number of features in the cue-fields of the vectors under comparison. The cosine measure of similarity is conceptually consistent with the similarity measure used in the Minerva 2 model. However, it normalizes over vector length and, thus, handles the continuous and added range of values of memory traces that result from the discrepancy encoding operation.

Second, we added noise to the echo by sampling a value from the range  $\pm 0.001$  to each of its elements. The change is pragmatic. Minerva 2 is a model of single-trial learning. If noise is not added to the echo, Minerva-AL learns too quickly, often in a single trial. Thus, on a first learning trial, the echo contains noise alone.

In the remainder of the work presented here, we apply Minerva-AL against associative learning protocols. All of the simulations will follow the same general method. Cues  $A$  through  $D$  are represented by separate 120 element vectors. For cue  $A$ , elements 1–20 will be set to one and all other elements equal to zero. For cue  $B$ , elements 21–40 will be set to one and all other elements equal to zero. For cue  $C$ , elements 41–60 will be set to one and all other elements equal to zero. For cue  $D$ , elements 61–80 will be set to one and all other elements equal to zero. Features of the experimental context (i.e., lighting, odors, ambient noise) are represented using a 120-element vector with elements 81–100 set to one and all other elements set to zero. Finally, outcome  $X$  will be a 120-element vector with elements 101–120 set to one and all other elements equal to zero.<sup>1</sup>

On each learning trial, a probe is constructed and then applied to memory; although we will not include a notation for the context in the probe, the context will be a part of every probe on every trial. The information retrieved into the echo is used to measure learning and to shape encoding via the discrepancy encoding operation. Learning is denoted by a conditional. For example,  $X|B$  refers to “retrieval of  $X$  given  $B$ ,” and  $X|AB$  refers to “retrieval of  $X$  given  $AB$ .” Learning in the model ranges between  $-1$  and  $+1$ . Positive values signal that the presented cue or cues retrieved  $X$  from memory (i.e., that the presented cues caused an expectation that  $X$  will follow). Negative values signal that the presented cue or cues retrieved an inverse representation of  $X$  from memory (i.e., that the cues caused an expectation that  $X$  will *not* follow). A value of zero signals that the presented cues elicited no expectation for  $X$ .

Despite differences between Minerva 2 and Minerva-AL, Minerva-AL preserves the spirit of its parent theory. Each learning trial is recorded in memory as a unique trace. At retrieval, the probe contacts all traces in parallel, and a weighted sum of the information in memory is retrieved (i.e., the echo). Information retrieved by a cue is quantified from the echo, with learning measured as the cue's ability to retrieve the target outcome from memory.

### Simulations of the formal model

We now simulate a series of associative learning protocols. In Section 1, we simulate acquisition and extinction. In Section 2, we simulate variants of inhibition. In Section 3, we simulate cue interaction protocols. In Section 4, we

<sup>1</sup> In a previous article, the elements of a representation could take values  $+1$  and  $-1$ . Assigning  $+1$  to the representations is a simplifying assumption and does not change the model or its behavior.

simulate discrimination and generalization. In Section 5, we simulate retrospective reevaluation.

## Section 1: Acquisition

*Acquisition and extinction* In a simple associative learning protocol, a cue,  $A$ , is presented, followed by an outcome,  $X$ . After several pairings, the cue elicits anticipation of  $X$ , a result called *acquisition*. If  $A$  is then presented alone, it ceases to elicit anticipation of  $X$ , a result called *extinction*.

We applied Minerva-AL to a 200-trial acquisition/extinction protocol. Trials 1–100 were acquisition trials (i.e.,  $A$  followed by  $X$ ); trials 101–200 were extinction trials (i.e.,  $A$  presented alone). On each trial a trace was stored to memory. On acquisition trials, the event vector,  $E$ , was equal to  $A + X$ . On extinction trials,  $E = A$ . On each trial, we evaluated learning by probing memory with  $A$  and indexing learning as retrieval of  $X$  given  $A$ .

We conducted 25 independent replications of the protocol for each of three values of  $L = .33, .67, \text{ and } 1.0$ . Figure 1 presents the results of the simulation.

As is shown in Fig. 1, Minerva-AL anticipates negatively accelerated acquisition and extinction curves. The rate of learning was correlated with the encoding parameter,  $L$ .

How does Minerva-AL explain acquisition and extinction? At the outset of a simulation, memory is empty (a matrix of 0 s). Consequently, on trial 1,  $A$  retrieves only noise into the echo. Because the echo contains only noise, retrieval of  $X$  given  $A$  is approximately zero, and the event vector is encoded strongly to memory (see Eq. 6). On trial 2,  $A$  retrieves the trace stored on trial 1. Because the trace records the pairing of  $A$  and  $X$  on trial 1, retrieval of  $X$  given  $A$  improves on trial 2. On trial 3, cue  $A$  now recruits the traces stored on trials 1 and 2. Summing both traces into the echo produces an even better representation of  $X$ . Recruiting each additional trace into the echo produces a systematic and cumulative benefit retrieval of  $X$  given  $A$  over successive trials.

On trial 1 of extinction,  $A$  retrieves a near perfect representation of  $X$ . However,  $X$  is not presented. Because Minerva-AL anticipates  $X$  but  $X$  is not presented, the trace stored to memory codes for a positive representation of  $A$  paired with a negative or inverse representation of  $X$  (i.e.,  $-X$ ). On trial 2 of extinction,  $A$  retrieves the traces from the acquisition phase, as well as the first trace from the extinction phase (i.e., the trace that records a negative representation of  $X$ ). Because the negative representation of  $X$  is retrieved, retrieval of  $X$  given  $A$  is diminished. This process repeats over the extinction trials and results in a cumulative impairment of  $X|A$ , until  $X|A$  is approximately zero. Critically, Minerva-AL does not expect that extinction will make  $A$  into a conditioned inhibitor.

The simulation illustrates that Minerva-AL can learn to anticipate an outcome given a cue and why anticipation of the outcome diminishes with presentation of the cue alone. We now turn to a description of reacquisition.

*Rapid reacquisition* Relearning is faster than new learning, a phenomenon called *rapid reacquisition* (Ebbinghaus, 1885/1964). Rapid reacquisition is assessed in a three-phase procedure (see Table 1). Phases 1 and 2 are acquisition and extinction, respectively. Phase 3 is the reacquisition phase where the contingency from phase 1 is re-presented. Rapid reacquisition is observed when relearning of an already acquired contingency is more efficient than learning of a novel contingency.

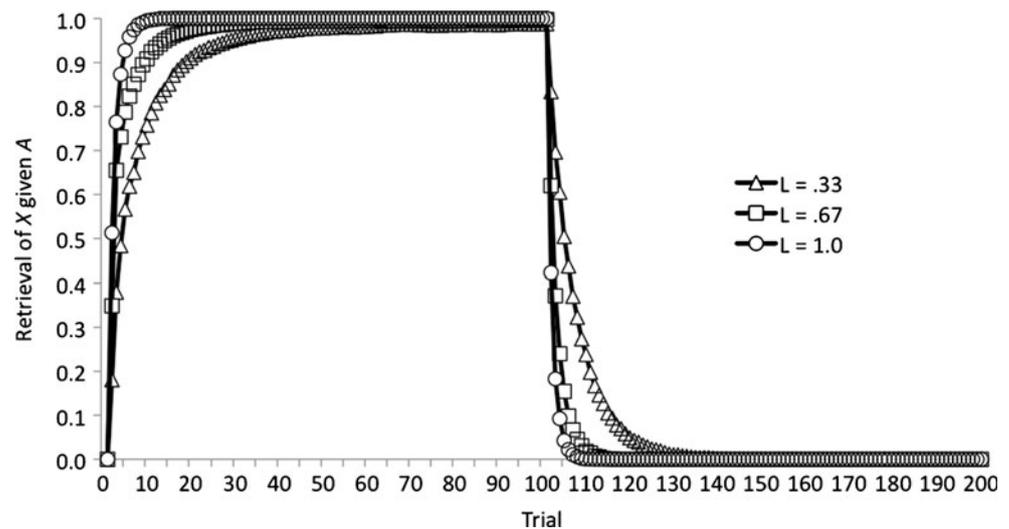
We applied Minerva-AL to the three-phase protocol in Table 1. The last three columns in Table 1 report the mean number of trials that Minerva-AL needed to achieve a criterion of  $X|A \geq .95$  in the reacquisition condition (and to achieve  $X|B \geq .95$  in the control condition). As is shown, Minerva-AL predicts rapid reacquisition. The model took fewer trials to relearn the old  $X|A$  contingency than it did to learn the novel  $X|B$  contingency.

An explanation of reacquisition picks up where our explanation of acquisition and extinction left off. On trial 1 of reacquisition,  $A$  retrieves a representation of  $X$  that is a weighted sum of traces from the acquisition and extinction phases. Because memory included traces from both the acquisition and extinction trials (i.e.,  $A \rightarrow X$  and  $A \rightarrow \neg X$  traces, respectively), the echo reflected uncertainty over whether  $X$  ought to be expected. Consequently, the re-presentation of the  $A \rightarrow X$  contingency was surprising to the model. Because memory highlights unexpected over expected events, Minerva-AL strongly encoded the  $A \rightarrow X$  contingency to memory, thus accelerating relearning.<sup>2</sup>

*Summary* Whereas acquisition and extinction do not challenge any existing theories of learning, they are a necessary benchmark for any theory of associative learning. In Section 2, we apply Minerva-AL to problems of inhibition.

<sup>2</sup> We have presented averaged learning curves. Like averaged empirical learning curves, the model anticipates a gradual and negatively accelerated growth in  $A$ 's ability to retrieve  $X$ . However, query Gallistel, Balsam, and Fairhurst (2004), for example, have argued that averaging over learners misrepresents the learning of individuals within the group average. Namely, individual's learning is best described as a step function. The difference is an important one. To explore the issue, we applied a decision rule for responding where response strength is a logistic transformation of how well  $X$  is retrieved by  $A$ . Using that transformation gave learning curves from individual simulations a more steplike shape. Averaging over the individual curves produced an averaged negatively accelerated curve, like the one depicted in Fig. 1.

**Fig. 1** Acquisition (trials 1–100) and extinction (trials 101–200). Means and standard errors computed across 25 replications of the procedure



## Section 2: Inhibition

Since Pavlov (1927), researchers have examined the construct of conditioned inhibition: an acquired tendency of a cue to attenuate responding to an outcome. In this section, we demonstrate that Minerva-AL predicts conditioned inhibition in direct, summation, and retardation tests.

*Direct test for conditioned inhibition* Table 2 describes a direct test for conditioned inhibition. The protocol includes a training phase followed by a test. The training phase mixes two types of trials: trials where  $A$  is presented followed by  $X$  and trials where  $AB$  is presented without  $X$ . At test, retrieval of  $X$  is tested given both  $A$  and  $B$  as retrieval cues. Conditioned inhibition is observed if  $X|B$  is less than zero.

We simulated the procedure in Table 2. As is shown, Minerva-AL retrieved a negative representation of  $X$  when probed with  $B$  and a positive representation of  $X$  when probed with  $A$ . Minerva-AL predicts conditioned inhibition.

In Minerva-AL, conditioned inhibition follows from a series of covert events in memory. On the  $A \rightarrow X$  trials, Minerva-AL learns that  $A$  predicts  $X$ . Consequently, Minerva-AL anticipates  $X$  on  $AB$  trials. Because  $X$  is not presented on  $AB$  trials, the trial-specific trace records a combination of  $+A$ ,  $+B$ , and  $-X$ . At test,  $B$  retrieves the  $+A$ ,  $+B$ , and  $-X$  traces encoded from the  $AB$  trials to retrieve  $-X$  into the echo.

*Summation test for conditioned inhibition* Whereas a direct test for conditioned inhibition is sufficient in a model

analysis, it is problematic for empirical work (see Cole, Barnet, & Miller, 1997; Rescorla, 1969). To circumvent the problem, researchers demonstrate conditioned inhibition using a summation test.

Table 3 describes a summation test for conditioned inhibition. Phase 1 restates the conditioned inhibition protocol (i.e., intermixed  $A \rightarrow X$  trials and  $AB$  trials). Phase 2 presents a novel cue,  $C$ , followed by  $X$ . After, retrieval of  $X$  given  $BC$ ,  $CD$ , and  $C$  is tested. According to the logic of the summation test, if  $B$  is a conditioned inhibitor of  $X$ , then retrieval  $X$  given  $BC$  ought to be worse than retrieval  $X$  given  $C$  alone (e.g., Rescorla, 1969, 1971).

We conducted 25 independent replications of the experimental and control procedures in Table 3. We report mean conditional recall of  $X$  given  $BC$ ,  $CD$ , and  $C$  as cues. As is shown, Minerva-AL predicts the summation test for conditioned inhibition:  $X|BC < X|C$ . Adding  $B$  to the probe (i.e., the conditioned inhibitor) weakens retrieval of  $X$ . Of course, the difference in  $X|BC$  and  $X|C$  might reflect the difference in the number of retrieval cues (two versus one, respectively). To rule out the confound, we compared  $X|BC$  with  $X|CD$ , where  $D$  is a neutral cue and  $B$  is a conditioned inhibitor. The simulation shows that  $X|BC$  was worse than  $X|CD$ . The disadvantage for  $X|BC$  confirms that the disadvantage for  $X|BC$  relative to  $X|C$  reflects more than a simple difference in the number of cues in the test probe.

**Table 1** Simulation of rapid reacquisition

Condition	Training			$L$		
	Phase 1	Phase 2	Test	.33	.67	1.00
Reacquisition	50 $A \rightarrow X$	50 $A$	$A \rightarrow X$	36.2 (5.9)	19.7 (5.1)	5.1 (0.2)
Control	50 $A \rightarrow X$	50 $A$	$B \rightarrow X$	62.2 (7.4)	32.1 (6.4)	3.8 (0.1)

Note. Means and standard errors computed over 25 independent replications. Standard errors are in parentheses.

**Table 2** Simulation of the direct test for conditioned inhibition

Training	Test	$L$		
		.33	.67	1.00
50 $A \rightarrow X$ / 50 $AB$	$X A$	.93 (.02)	.97 (.01)	.99 (.01)
	$X B$	-.16 (.01)	-.21 (.02)	-.32 (.04)

Note. Means and standard errors computed over 25 independent replications. Standard errors are in parentheses.

**Retardation test for conditioned inhibition** When a summation test is not possible, researchers use a retardation procedure to test condition inhibition (described in Table 4). Like the summation test, the training phase of the retardation test restates the conditioned inhibition protocol (i.e., intermixed  $A \rightarrow X$  trials and  $AB$  trials). In contrast to the summation test, conditioned inhibition to  $B$  is assessed by measuring the number of trials that the subject needs to learn  $X|BC$  to a specified criterion. If cue  $B$  is a conditioned inhibitor, it ought to take more trials to learn  $X|BC$  than  $X|C$ .

We conducted 25 independent replications of the procedures in Table 4. As is shown, Minerva-AL needed more trials to achieve  $X|BC \geq .95$  than it needed to achieve  $X|C \geq .95$  after training with intermixed  $A \rightarrow X$  and  $AB$  trials (i.e., experimental treatment). The difference confirms that including  $B$  as part of the retrieval cue at test retards retrieval of  $X$ .

Minerva-AL predicts conditioned inhibition in the direct, summation, and retardation protocols for conditioned inhibition. All three simulations demonstrate that Minerva-AL can learn to both anticipate  $X$  given a cue and to anticipate  $\neg X$  (i.e., not  $X$ ) given a cue. The next two simulations show that Minerva-AL handles inhibition in two other protocols.

**External inhibition** External inhibition refers to the situation where adding a novel cue to a test probe weakens the expression of an established association. Table 5 summarizes the external inhibition protocol. In a training phase,  $A$  is presented followed by  $X$  (i.e., simple acquisition). Afterward, retrieval of  $X|A$  and retrieval of  $X|AB$  are tested. In this case,  $B$  is a neutral untrained cue. External inhibition is observed when retrieval of  $X|A$  is greater than retrieval of  $X|AB$ .

As is shown in Table 5, Minerva-AL anticipates external inhibition:  $X|A > X|AB$ . The orthodox explanation of external inhibition is that adding a novel cue to the test probe forces a generalization decrement (e.g., Pavlov, 1927). Minerva-AL's explanation of the effect is slightly different. Presenting  $AB$  causes a different activation of traces in memory than does  $A$  alone. The difference in trace activations drives the differences in Table 5.

**Latent inhibition** Latent inhibition is observed when preexposure to a cue slows subsequent learning with it (Mackintosh, 1975; Rescorla, 1971). We simulated a simple latent inhibition protocol with a preexposure phase followed by a test phase (see Table 6). In the preexposure phase,  $A$  is

presented alone (i.e., without consequence). At test,  $A$  is presented followed by  $X$ . At test, learning of the  $X|A$  contingency is measured by the logic of a retardation test. Latent inhibition is observed if acquisition for  $X|A$  is slower than in a control condition excluding the preexposure phase.

We conducted 25 independent replications of the latent inhibition procedure and measured the number of trials that Minerva-AL took in the test phase until  $X|A \geq .95$ . As is shown in Table 6, the number of trials to reach  $X|A \geq .95$  increased with the number of preexposure trials. The result is the latent inhibition effect.

Minerva-AL explains latent inhibition as a consequence of proactive interference. Each preexposure trial leaves a trace in memory:  $A$  paired with noise. At test,  $A$  retrieves those traces, including the noise that interferes with the resolution of  $X$  in the echo. Later, when training  $A \rightarrow X$ , the noise in the preexposure traces is weighed out of the echo.

Hall and Channell (1985) demonstrated that latent inhibition is context specific by preexposing  $A$  in one context and then testing acquisition of  $A \rightarrow X$  in another (see also Nelson & Sanjuan, 2006). We resimulated the latent inhibition protocol from Table 6, but, in the preexposure phase,  $A$  was presented in one context, and in the experimental phase, the  $A \rightarrow X$  contingency was presented in a different context (i.e., this was accomplished by setting the usual context field included in the probes for phase 1 to zero and adding a new context vector). The rate at which Minerva-AL acquired the  $X|A$  contingency was substantially less affected by the preexposure manipulation. We conclude that Minerva-AL acknowledges the context-specific nature of latent inhibition.

**Hall–Pearce negative transfer** Hall and Pearce (1979) generated an example of latent inhibition where, in the preexposure phase of the protocol,  $A$  is presented, followed by a muted version of  $X$ —rather than  $A$  being presented alone (see also Ayres, Moore, & Vigorito, 1984). We simulated this procedure by replicating the latent inhibition protocol in Table 6, but instead of presenting  $A$  alone during the preexposure phase, we presented the model with  $A$  followed by a muted version of  $X$ . To mute  $X$ , we multiplied its representation by a weight of 0.1, thus rewriting each of the 1.0 values in the vector for  $X$  as values of 0.1. In the second phase of the protocol, we presented  $X$  at full strength (i.e., as we did in the preceding two simulations). The simulation reproduced the Hall and Pearce effect. Namely, following preexposure to the  $A$  followed by  $X(.1)$ , it took more trials to reach  $X|A \geq .95$  in phase 2 of the procedure than following preexposure to  $A$  followed by  $X(1.0)$ . The difference held over all values of  $L$ .<sup>3</sup>

<sup>3</sup> Savastano, Yin, Barnet, and Miller (1998) have demonstrated that the Hall–Pearce negative transfer result weakens when the timing of the cue–outcome pairing is varied at training at test. Minerva-AL has no mechanism for modeling the timing of cues and outcomes and, so, is unable to model this aspect of the result.

**Table 3** Simulation of summation test for conditioned inhibition

Condition	Training		Test	<i>L</i>		
	Phase 1	Phase 2		.33	.67	1.00
Summation	50 $A \rightarrow X$ / 50 $AB$	50 $C \rightarrow X$	$X BC$	.13 (.02)	.18 (.02)	.05 (.07)
Control (1)	50 $A \rightarrow X$ / 50 $AB$	50 $C \rightarrow X$	$X CD$	.93 (.01)	.92 (.01)	.89 (.02)
Control (2)	50 $A \rightarrow X$ / 50 $AB$	50 $C \rightarrow X$	$X C$	.99 (.00)	.99 (.00)	1.0 (.00)

Note. Means and standard errors computed over 25 independent replications. Standard errors are in parentheses.

**Summary** Minerva-AL captures conditioned inhibition in the direct, summation, and retardation protocols. It captures external inhibition and latent inhibition. It captures the Hall–Pearce negative transfer effect.

### Section 3: Cue competition

A hallmark of associative learning is that concurrently presented cues compete for associative strength. The principle is hardwired into the learning rules of many models (e.g., Rescorla–Wagner) but is not hardwired into Minerva-AL. In this section, we show that Minerva-AL appreciates the complications of cue competition effects in learning.

**Blocking** Blocking is the classic demonstration of cue competition. A blocking protocol includes two successive training phases followed by a test phase (Kamin, 1969; Rescorla & Wagner, 1972). In phase 1, the cue,  $A$ , is presented, followed by an outcome,  $X$ . In phase 2, the cue compound,  $AB$ , is presented, followed by  $X$ . Following training, retrieval of  $X|B$  is tested. Table 7 summarizes the procedure and its control conditions. Blocking is observed if retrieval of  $X$  given  $B$  is weakest in the blocking condition.

We simulated the blocking and control procedures in Table 7. As is shown, Minerva-AL anticipated the blocking effect:  $X|B$  was smaller in the blocking condition than in either of the control conditions. The size of the blocking effect (i.e., the difference in  $X|B$  between the blocking and control conditions) correlates with  $L$ .

Minerva-AL's explanation for blocking is clear. In phase 1,  $A$  is established as a retrieval cue for  $X$ . In phase 2,  $AB$  retrieves  $X$ . Because  $AB$  retrieves  $X$ , the biggest discrepancy between the echo and the event vector,  $E = A + B + X$ , is the presence of  $B$ . Consequently, the trace stored includes a

strong representation of  $B$  paired with a weak representation of  $X$ . Consequently, at test,  $B$  retrieves a weak representation of  $X$ . The simulation presents an explanation of blocking that follows from the storage and retrieval of instances from memory.

**Overshadowing** In an overshadowing experiment, two cues,  $A$  and  $B$ , are presented together, followed by an outcome,  $X$ . Typically, one cue,  $B$ , is made more salient (e.g., louder, brighter) than the other. Overshadowing is observed if the strength of conditioning for the less salient cue  $A$  is attenuated when trained in compound with  $B$ . In demonstrations, the degree of attenuation is proportional to the relative saliency of the damped and undamped cues (Kamin, 1969; Mackintosh, 1971).

To model overshadowing, we represented cue saliency by introducing saliency weights,  $\alpha$ . To illustrate, if cue  $A = [1.0, 1.0, 1.0, 1.0]$  and  $\alpha_A = 0.1$ , then  $\alpha_A \times A = [0.1, +0.1, 0.1, +0.1]$ . We constrain saliency weights between 0.0 and 1.0; the smaller  $\alpha$  is, the less salient the cue. The manipulation of cue saliency is identical to our previous manipulation of outcome saliency to simulate the Hall–Pearce negative transfer effect.

We simulated overshadowing using a training/test protocol. In the training phase,  $AB$  was presented followed by  $X$ . We varied the cue saliency of  $A$  (i.e.,  $\alpha_A = 0.1, 0.6$ , and  $1.0$ ) while holding the cue saliency of  $B$  constant (i.e.,  $\alpha_B = 1.0$ ). The procedure is summarized in Table 8. If Minerva-AL predicts overshadowing,  $X|A$  ought to be greater in the control condition where  $A$  is trained independently of  $B$  than in the three experimental conditions where  $A$  is trained with  $B$ . Moreover, the strength of overshadowing ought to increase as the disparity between the saliency of  $A$  and  $B$  is increased.

**Table 4** Simulation of retardation test for conditioned inhibition

Condition	Training	Test	<i>L</i>		
			.33	.67	1.00
Experimental	50 $A \rightarrow X$ / 50 $AB$	$BC \rightarrow X$	46.6 (3.7)	16.3 (0.9)	5.6 (0.7)
Control		$C \rightarrow X$	29.4 (1.5)	13.2 (1.0)	4.2 (0.2)

Note. Means and standard errors computed over 25 independent replications. Standard errors are in parentheses.

**Table 5** Simulation of external inhibition

Condition	Training	Test	<i>L</i>		
			.33	.67	1.00
External inhibition	50 $A \rightarrow X$	$X AB$	.80 (.02)	.85 (.02)	.90 (.02)
Control	50 $A \rightarrow X$	$X A$	.98 (.00)	.99 (.00)	1.0 (.00)

*Note.* Means and standard errors computed over 25 independent replications. Standard errors are in parentheses.

Table 8 shows the results of the simulations as a function of  $L$  and  $\alpha_A$ . Consistent with the overshadowing effect,  $X|A$  was weaker when  $A$  was trained in compound with  $B$ , and the size of the overshadowing effect increased with the difference in saliency between cues  $A$  and  $B$ .

**Overexpectation** An overexpectation effect is observed when responding to a pair of well-established cues is diminished after they have then been given additional compound training (e.g., Kremer, 1978). Table 9 summarizes the protocol. In phase 1 of training,  $A \rightarrow X$  and  $B \rightarrow X$  trials are intermixed. In phase 2 of training,  $AB$  is presented followed by  $X$ . Overexpectation is observed if retrieval of  $X|A$  is worse following  $AB \rightarrow X$  versus  $AC \rightarrow X$  training or following no additional training in phase 2. The overexpectation effect is of interest here because it escapes any straightforward rational explanation: It is not immediately clear why cue  $A$  might be made a worse predictor of  $X$  as a consequence of additional exposure to the  $AB \rightarrow X$  contingency.

Table 9 presents the results of the simulation. As is shown, Minerva-AL predicts the overexpectation effect:  $X|A$  is smaller in the overexpectation condition (i.e., following the  $AB \rightarrow X$  training in phase 2) than in either of the control conditions. The size of the overexpectation effect correlates with  $L$ .

Minerva-AL explains overexpectation to be a natural outcome of operations in memory. In phase 1, the model learns the  $A \rightarrow X$  and  $B \rightarrow X$  contingencies. In phase 2,  $AB$  retrieves an especially clear representation of  $X$ . Because the representation of  $X$  is especially clear, discrepancy encoding records that  $A$  and  $B$  were paired with what would seem to be a weakened or less salient presentation of  $X$ . At test, these traces weaken the representation of  $X$  that is

retrieved into the echo. That is, overexpectation of  $X$  in phase 2 makes  $A$  a worse retrieval cue for  $X$ .

**Superconditioning** Superconditioning is a corollary of overexpectation. In superconditioning, pairing a novel cue with a conditioned inhibitor benefits, rather than impedes, learning of a novel cue (Williams & McDevitt, 2002). The effect occurs because the conditioned inhibitor retrieves an expectation that the outcome will not occur. Consequently, when the outcome does occur it is “super surprising.”

Table 10 shows a superconditioning procedure. Phase 1 intermixes  $A \rightarrow X$  and  $AB$  trials. In phase 2,  $BC$  is presented followed by  $X$ , or in a control condition,  $CD$  is presented followed by  $X$ . Superconditioning is observed when  $X|C$  is greater following  $BC \rightarrow X$  training in phase 2 of the procedure than following  $CD \rightarrow X$  training in phase 2 of the procedure.

Simulation results are presented in the final three columns of Table 10. As is shown, learning of the  $X|C$  contingency was better in the superconditioning condition (i.e., after  $C$  had been paired with a conditioned inhibitor,  $B$ , in phase 2 of training), as compared with the control condition (i.e., after  $C$  had been paired with a neutral cue,  $D$ , in phase 2 of training). Minerva-AL anticipates the superconditioning effect.

Minerva-AL’s explanation for superconditioning highlights the discrepancy encoding operation. In phase 2 of training,  $B$  retrieved an inverse representation of  $X$  into the echo. Consequently, presentation of  $X$  was more surprising to the model than if it had no expectations at all. The violation of expectation resulted in strong encoding of the unexpected events:  $C$  because it was a novel event and  $X$  because the model expected it to not occur. Because  $C$  and  $X$  were both unexpected, they were strongly encoded, thereby accelerating learning of the  $X|C$  contingency.

**Table 6** Simulation of latent inhibition

Condition	Training	Test	<i>L</i>		
			.33	.67	1.00
0 pretrials		$A \rightarrow X$	32.1 (1.7)	13.4 (1.0)	6.3 (0.5)
10 pretrials	10 $A$	$A \rightarrow X$	54.2 (3.0)	31.2 (3.3)	6.2 (0.5)
30 pretrials	30 $A$	$A \rightarrow X$	61.6 (4.9)	35.8 (5.6)	6.4 (0.3)

*Note.* Means and standard errors computed over 25 independent replications. Standard errors are in parentheses.

**Table 7** Simulation of blocking

Condition	Training		Test	<i>L</i>		
	Phase 1	Phase 2		.33	.67	1.00
Blocking	50 $A \rightarrow X$	50 $AB \rightarrow X$	$X B$	.32 (.01)	.30 (.01)	.26 (.02)
Control (1)	50 $C \rightarrow X$	50 $AB \rightarrow X$	$X B$	.65 (.01)	.76 (.01)	.89 (.01)
Control (2)		50 $AB \rightarrow X$	$X B$	.66 (.02)	.75 (.01)	.85 (.01)

*Note.* Means and standard errors computed over 25 independent replications. Standard errors are in parentheses.

**Summary** Minerva-AL models cue competition effects. The ability is clear even though the model does not include an explicit rule for how individual cues ought to be handled when presented individually and in compound. In the section on discrimination that follows, we elaborate on this point to show that, in spite of no explicit rule for handling single cues and compounds, Minerva-AL can keep track of its expectations in a way that matches the behavior of real learners.

#### Section 4: Discrimination and generalization

Discrimination is an ability to respond differently to alternate albeit similar and related cues. The classic demonstration of discrimination involves presentation of one cue,  $A$ , followed by  $X$  and another cue,  $B$ , without  $X$ . Discrimination is observed when  $A$  elicits anticipation of  $X$  but  $B$  does not.

We applied Minerva-AL to the simple discrimination of nonoverlapping cues procedure. The simulation involved 50 blocks, each of which had two trials. One trial presented  $A$  followed by  $X$ . The other trial presented  $B$  (without  $X$ ). Following each block, we measured  $X|A$  and  $X|B$ .

Figure 2 shows  $X|A$  and  $X|B$  over the 50 blocks for  $L = .67$ . As is shown, early in the procedure, Minerva-AL expected that  $X$  would follow  $A$  or  $B$ . With additional training, the errant expectation was resolved, and discrimination was achieved by about block 20, where  $X|A \approx 1$  and  $X|B \approx 0$ . The trajectory of discrimination is consistent with data from laboratory experiments (e.g., Redhead & Pearce, 1995). We conclude that Minerva-AL handles simple discrimination of nonoverlapping cues.

The simulation shows that Minerva-AL discriminates reinforced from unreinforced cues. However, a competent model ought to be capable of discriminating cues that are presented singly (i.e., elemental cues) and in compound (i.e., configural cues). For example, the model should learn that  $A$  predicts  $X$  when paired with  $B$ , but not when it is presented alone. Moreover, and more critically, the model ought to learn similar nonlinear discriminations such as negative patterning. To better address discrimination learning, we applied Minerva-AL, first, to the linear positive-patterning problem and, second, to the nonlinear negative-patterning problem.

**Positive patterning** Positive patterning represents a linear discrimination. In positive patterning, the learner must respond to a cue compound,  $AB$ , without responding to its constituents (Pavlov, 1927; Wasserman & Miller, 1997; Young, Wasserman, Johnson, & Jones, 2000). The discrimination ensures that the model recognizes differing schedules to a cue presented alone and to that same cue presented in a compound.

We applied Minerva-AL to the positive-patterning procedure. In each of 50 successive blocks, we presented the model with intermixed trials of  $AB \rightarrow X$ ,  $A$ , and  $B$ . As is shown in Fig. 3, Minerva-AL learned the discrimination: It retrieved  $X$  given  $AB$  but did not retrieve  $X$  given  $A$  or  $X$  given  $B$ , and discrimination developed quickly and grew increasingly strong with practice.

Positive pattern discrimination is a linear discrimination problem. This means that the problem can be solved by responding if both  $A$  and  $B$  are present and not responding if only one is present. A more difficult nonlinear discrim-

**Table 8** Simulation of overshadowing

Condition	Training	Test	<i>L</i>		
			.33	.67	1.00
Control ( $\alpha_A=1.0$ )	50 $A \rightarrow X$	$X A$	.98 (.00)	1.0 (.00)	1.0 (.00)
Experimental ( $\alpha_A=1.0$ )	50 $AB \rightarrow X$	$X A$	.63 (.02)	.79 (.02)	.88 (.02)
Experimental ( $\alpha_A=0.6$ )	50 $AB \rightarrow X$	$X A$	.56 (.02)	.72 (.02)	.85 (.02)
Experimental ( $\alpha_A=0.1$ )	50 $AB \rightarrow X$	$X A$	.50 (.03)	.63 (.03)	.82 (.02)

*Note.* Means and standard errors computed over 25 independent replications. Standard errors are in parentheses.

**Table 9** Simulation of overexpectation

Condition	Training		Test	<i>L</i>		
	Phase 1	Phase 2		.33	.67	1.00
Overexpectation	25 $A \rightarrow X / 25 B \rightarrow X$	50 $AB \rightarrow X$	$X A$	.80 (.01)	.80 (.02)	.72 (.04)
Control (1)	25 $A \rightarrow X / 25 B \rightarrow X$	50 $AC \rightarrow X$	$X A$	.96 (.01)	.98 (.01)	.99 (.00)
Control (2)	25 $A \rightarrow X / 25 B \rightarrow X$		$X A$	.98 (.00)	1.0 (.00)	1.0 (.00)

*Note.* Means and standard errors computed over 25 independent replications. Standard errors are in parentheses.

ination problem is negative patterning that requires the learner to discriminate response patterns in a more interesting fashion.

**Negative patterning** Negative patterning is the opposite of positive patterning. In the protocol, a learner must anticipate  $X$  following presentation of  $A$  or  $B$  in isolation without anticipating  $X$  following presentation of  $A$  and  $B$  in compound (e.g., Pearce, 1994; Wasserman & Miller, 1997; Young et al., 2000). The protocol is important for model evaluation because it requires the learner to compute a nonlinear discrimination—a feat that not all models are capable of (see Miller et al., 1995).

Figure 4 shows Minerva-AL's performance in a negative-patterning protocol that presented one example each of an  $A \rightarrow X$ ,  $B \rightarrow X$ , and  $AB$  contingency in each of 50 successive three-trial blocks. As is shown, Minerva-AL learned to anticipate  $X|A$  and  $X|B$  without anticipating  $X|AB$ . The ability suggests that Minerva-AL learned to respond selectively to the situations where  $A$  and  $B$  were presented in isolation.

Negative patterning is important because a simple elemental model of conditioning that assumes a simple relationship between stimuli presented singly and in compound will fail to predict the result. The failure follows because such models treat the associative strength of a compound to be a sum of associative strengths to the constituent elements. Thus, the response to  $AB$  ought always to be greater than the responses to  $A$  and  $B$  alone. As such, the linear discrimination of positive patterning is acknowledged to be easier to learn than the nonlinear discrimination of negative patterning (e.g., Bellingham, Gillette-Bellingham, & Kehoe, 1985; Rescorla, 1973; Young et al., 2000). We note that Minerva-AL anticipates

easier learning of the positive- than of the negative-patterning problems (see Figs. 3 and 4).

**Biconditional discrimination** In a biconditional discrimination protocol, the learner is required to learn four contingencies:  $AB \rightarrow X$ ,  $CD \rightarrow X$ ,  $BC \rightarrow \neg X$ , and  $AD \rightarrow \neg X$ . Critically, the four individual cues— $A$ ,  $B$ ,  $C$ , and  $D$ —are reinforced and unreinforced equally often over the course of the experiment. Thus, to learn the discrimination, the subject must learn the relationships between the cue compounds and the outcome. Learning theories that assume associative learning from stimulus elements to outcomes with the associative strength of a compound equal to the sum of associative strengths of its elements cannot accommodate the result (Lober & Lachnit, 2002).

We applied Minerva-AL to the biconditional discrimination problem. Figure 5 presents the simulation results. Following exposure to the contingencies, Minerva-AL learned the contingencies:  $X|AB$  and  $X|CD$  were both equal and greater than both  $X|BC$  and  $X|AD$ .

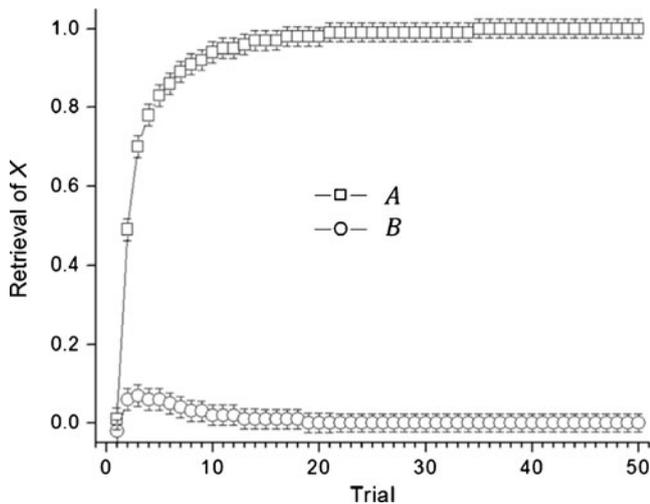
Animals learn the negative discrimination faster than the biconditional problem (Harris, Livesey, Gharaei, & Westbrook, 2008; although Harris & Livesey, 2008, show that the difference did not hold with humans). Inspection of the results with the negative patterning and biconditional discriminations shows that Minerva-AL mispredicts the learning advantage for negative over biconditional pattern discriminations (see Figs. 4 and 5).

**Discrimination of cues presented singly and in cue compounds** Thus far, we have simulated protocols—not data. In the next two simulations, we show not only that Minerva-AL accomplishes nonlinear discrimination, but also that it can match the details of empirical data. As a test base, we use two procedures from Redhead and

**Table 10** Simulation of superconditioning

Condition	Training		Test	<i>L</i>		
	Phase 1	Phase 2		.33	.67	1.00
Super-conditioning	50 $A \rightarrow X / 50 AB$	50 $BC \rightarrow X$	$X C$	.81 (.02)	.88 (.01)	.95 (.01)
Control	50 $A \rightarrow X / 50 AB$	50 $CD \rightarrow X$	$X C$	.67 (.01)	.77 (.02)	.85 (.02)

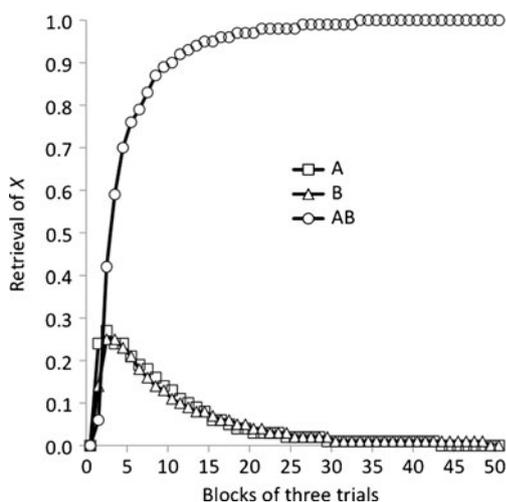
*Note.* Means and standard errors computed over 25 independent replications. Standard errors are in parentheses.



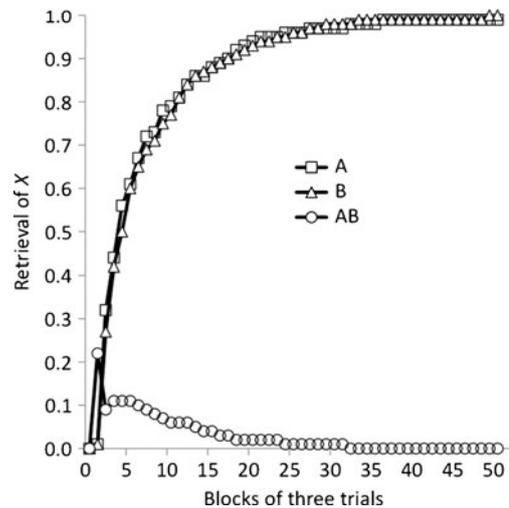
**Fig. 2** Simulation of simple discrimination learning:  $A \rightarrow X$  and  $B \rightarrow \neg X$ . Means and standard errors computed across performance in 25 replications of the procedure

Pearce's (1995) work. The first procedure had a preexposure phase that was followed by a test phase. The preexposure phase included 48 trials in which  $A$  was presented 8 times,  $AB$  was presented 8 times, and  $ABC$  was presented 32 times. The test phase that followed had 10 blocks. Each block included 48 trials in which  $A$  was presented 8 times followed by  $X$ ,  $AB$  was presented 8 times followed by  $X$ , and  $ABC$  was presented alone 32 times (i.e., without  $X$ ). Redhead and Pearce's results are reproduced in the inset of Fig. 6. As is shown,  $X|ABC < X|AB < X|A$ .

We simulated Redhead and Pearce's (1995) procedure for various levels of  $L$ . The results of the simulations are presented in the main panel of Fig. 6 for  $L = .1$ . As is shown, Minerva-AL reproduced the main features of Redhead and Pearce's data. First, consistent with Redhead



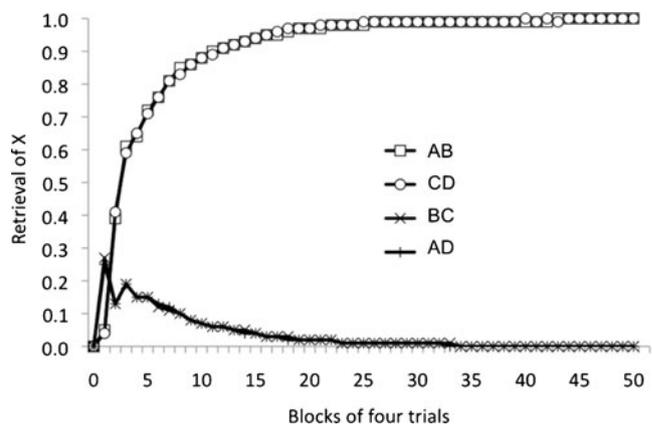
**Fig. 3** Simulation of positive patterning. Means and standard errors computed across performance in 25 replications of the procedure,  $L = .67$



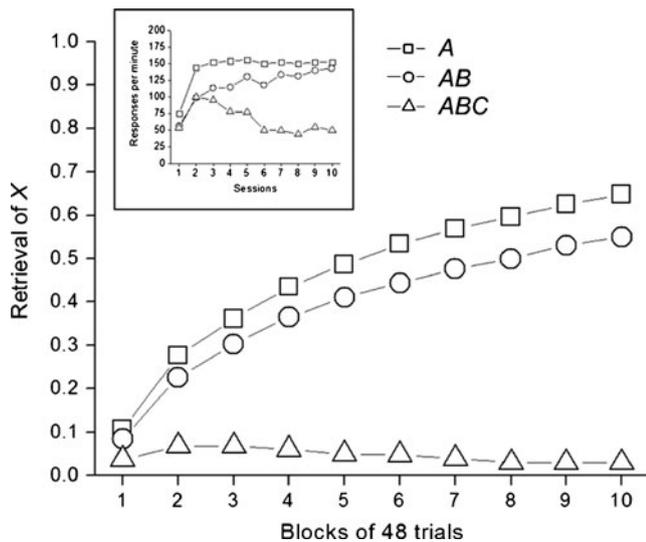
**Fig. 4** Simulation of negative patterning. Means and standard errors computed across performance in 25 replications of the procedure,  $L = .67$

and Pearce's data,  $X|A > X|AB$ . Second,  $X|ABC = 0$  (i.e.,  $ABC$  did not retrieve  $X$ ). The simulation shows that Minerva-AL can discriminate cues when presented separately from when they are presented in compound with either one or two others.

In a second experiment, Redhead and Pearce (1995) increased the number of discriminations presented concurrently. A preexposure phase had 54 trials that presented  $A$ ,  $B$ ,  $C$ ,  $AB$ ,  $AC$ , and  $BC$  3 times each and  $ABC$  36 times. The test phase that followed had four blocks. Each block included 54 trials. In each block,  $A$ ,  $B$ ,  $C$ ,  $AB$ ,  $AC$ , and  $BC$  were each presented 3 times followed by  $X$  and the compound  $ABC$  was presented 36 times without  $X$ . Redhead and Pearce summarized their data by comparing learning of single cues, cues presented in pairs, and the three cues presented in a complete compound (see the inset of Fig. 7). Responding was strongest to single cues, intermediate to cue pairs, and weakest to the full compound.



**Fig. 5** Simulation of biconditional discrimination. Means and standard errors computed across performance in 25 replications of the procedure,  $L = .67$



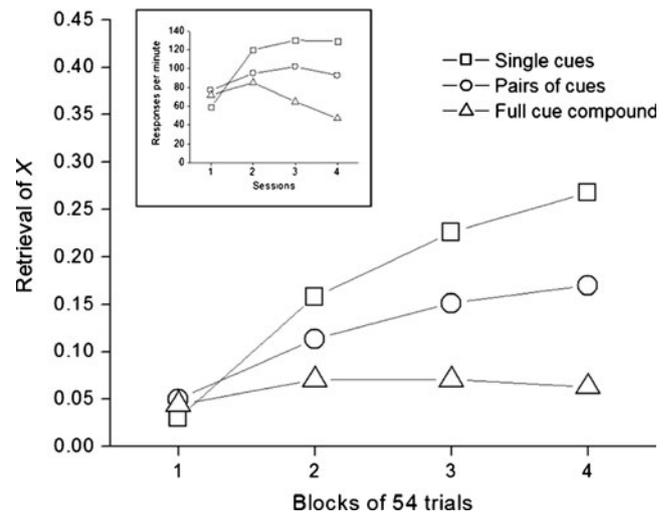
**Fig. 6** Simulation of Redhead and Pearce's (1995) Experiment 1. Means computed across performance in 25 replications of the procedure,  $L = .1$

We conducted 25 replications of the procedure for various levels of  $L$ . The results of the simulations are presented in Fig. 7 for  $L = .1$ . The model reproduced the major features of Redhead and Pearce's (1995) results. The discrimination between the individual cues,  $\{A, B, C\}$ , and the full compound  $\{ABC\}$  was acquired faster than the discrimination between pairs of cues  $\{AB, BC, AC\}$  and the full compound  $\{ABC\}$ . In addition, the single cues were the worst at retrieving  $X$  in block 1 but were the best cues for retrieving  $X$  in blocks 2, 3, and 4. This pattern of learning by the model matched the pattern in Redhead and Pearce's subjects. In previous work, Pearce (1994) applied his configural theory of learning to the procedures and generated approximately the same predictions as we have (compare the main panel in Fig. 7 with the right panel of Pearce's, 1994, Fig. 12).<sup>4</sup>

In Minerva-AL, traces record the events of individual trials. At retrieval, the similarity measure distinguishes the traces. For example, a probe that includes only  $A$  activates the traces in memory that code for  $A$  more strongly than traces that code for  $A$  in combination with  $B, C$ , and so on. By retrieving traces conjoining specific cues and outcomes, the theory exploits memory for individual trials and discriminates those trials in which a cue was presented alone and those in which it was presented in compound.

Minerva-AL's ability to model discrimination of cues when presented in various configurations follows from the

<sup>4</sup> We have fit Redhead and Pearce's (1995) data. Myers, Vogel, Shin, and Wagner (2001) have published data that contradict Redhead and Pearce. Namely, they found better discrimination of  $AB$  versus  $ABC$  than of  $A$  versus  $ABC$ . Thus, our fit of Redhead and Pearce still leaves Minerva-AL's ability to model discrimination with some degree of uncertainty.



**Fig. 7** Simulation of Redhead and Pearce's (1995) Experiment 2. Means computed across performance in 25 replications of the procedure,  $L = .1$

model's principle of instance-based representation. We now turn to a related problem of generalization.

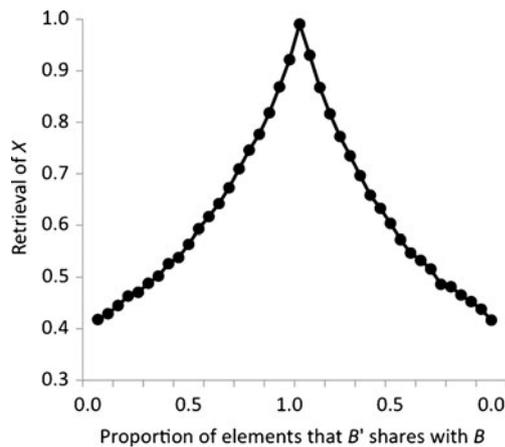
**Generalization** Generalization is the complement of discrimination. Generalization occurs when the response to a trained cue transfers to another, similar cue. For example, after it is learned that a tone of frequency  $f$  predicts an imminent shock, a tone of frequency  $f'$  will also elicit an avoidance response. Generally, the greater the similarity between the original trained cue,  $f$ , and its tested substitute,  $f'$ , the more complete is generalization (Moore, 1972).

We applied Minerva-AL to a simple generalization protocol. In a training phase,  $B$  was presented, followed by  $X$ . In a test phase, we evaluated the ability of probes that approximated  $B$  to retrieve  $X$ . The probes that approximated  $B$  were vectors of 120 elements where elements  $i$  through  $(i + 20)$  were set to 1 (where  $i$  is incremented between 1 through 20), with all other elements set to zero. Within the scheme for stimulus representation we have been using, these vectors are "in between cues  $A$  and  $C$ ." We used  $B$  as the training cue instead of  $A$  because it allowed us to model generalization to probes that overlapped early and late numbered features of  $B$ .

Figure 8 shows generalization to  $B'$  averaged over 25 independent replications where  $L = .15$ . As is shown, generalization (i.e., retrieval of  $X$ ) improved with the proportion of features that  $B'$  shared with  $B$ .

Minerva-AL's explanation of generalization shares its key premise with an explanation given by Pearce (1994). Namely, the proportion of elements that it shares with a previously established cue,  $B$ , determines the magnitude of generalization to the test probe,  $B'$ .

**Summary** We applied Minerva-AL to several discrimination problems. The model distinguishes the contingencies of a cue presented singly and in compounds. Minerva-AL



**Fig. 8** Simulation of stimulus generalization. Means and standard errors computed across performance in 25 replications of the procedure,  $L = .1$

accomplishes discrimination by selective retrieval of traces containing contingencies that match the test probe. This ability is a natural outcome for a model that treats associative learning as a corollary of the storage and retrieval of instances from memory.

#### Section 5: Retrospective revaluation

Most classical theories of learning assert that a cue must be present to acquire or lose associative strength with an outcome (e.g., Mackintosh, 1975; Pearce & Hall, 1980; Rescorla & Wagner, 1972; Wagner, 1981). However, demonstrations of retrospective revaluation contradict that assertion.

One example of retrospective revaluation is backward blocking. A backward blocking protocol has two successive training phases followed by a test. In phase 1 of training, a compound cue,  $AB$ , is presented, followed by an outcome,  $X$ . In phase 2 of training,  $A$  is paired with  $X$ . After phase 2 of training, retrieval of  $X$  given  $B$  is assessed. Backward blocking is observed if, at test, retrieval of  $X$  given  $B$  is worse than if phase 2 of training had not occurred. The procedure is shown in the top row of Table 11 (rows 2 and 3 in Table 11 show relevant control procedures).

At first blush, backward blocking implies symbolic inference: “I learned in phase one that the combination of

$A$  and  $B$  predicts  $X$ . But, in phase 2, I learned that  $A$  alone predicts  $X$ . Given the two contingencies, I can infer that, despite initial appearances,  $B$  did not predict  $X$  in phase 1 after all.” Whereas inferential reasoning might provide a sound explanation of the result with humans, inferential reasoning is a suspect explanation for learning in rats (Miller & Matute, 1996). Below, we explain retrospective revaluation without reasoned inference.

Van Hamme and Wasserman (1994) were the first to solve the problem of retrospective revaluation when they adapted the Rescorla–Wagner model to explain backward blocking. In the adapted model, phase 1 of training establishes  $A$  as a retrieval cue for  $B$  (i.e., a within-compound association). In phase 2,  $A$  activates  $B$  and, thus, opens  $B$  for learning. Van Hamme and Wasserman forced the backward-blocking result by applying a positive-learning rate to update the  $A \rightarrow X$  association and a negative-learning rate to update the  $B \rightarrow X$  association. Because the modified Rescorla–Wagner model does not include a process for inferential reasoning, it finesses the problem of attributing rationality to species in which such capabilities are suspect.

Whereas the thrust of Van Hamme and Wasserman’s (1994) explanation is sound, their computational solution was problematic because absent and presented cues had to be identified for the model and different learning parameters had to be applied to the two kinds. Witnauer and Miller (2011) have since resolved the problem and showed that the model handles retrospective revaluation, including second-order revaluation.

In the next simulation, we show that Minerva-AL also meets the challenge. Although we have simulated backward blocking, recovery from blocking, and backward conditioned inhibition elsewhere (see Jamieson, Hannah & Crumpt 2010), we re-present those simulations here. Afterward, we extend our previous analysis to include additional examples of retrospective revaluation including second-order revaluation.

*Backward blocking* We simulated a standard backward-blocking procedure. The procedure included two successive training phases, followed by a test. In training phase 1,  $AB$  was presented followed by  $X$ . In training phase 2,  $A$  was

**Table 11** Simulation of backward blocking

Condition	Training		Test	$L$		
	Phase 1	Phase 2		.33	.67	1.00
Backward blocking	50 $AB \rightarrow X$	50 $A \rightarrow X$	$X B$	.43 (.03)	.21 (.06)	.07 (.01)
Control (1)	50 $AB \rightarrow X$	50 $C \rightarrow X$	$X B$	.64 (.02)	.76 (.02)	.89 (.01)
Control (2)	50 $AB \rightarrow X$		$X B$	.60 (.02)	.78 (.02)	.90 (.01)

*Note.* Means and standard errors computed over 25 independent replications. Standard errors are in parentheses.

presented followed by  $X$ . Following phase 2, retrieval of  $X$  given  $B$  was tested. The design is shown in Table 11 along with two control conditions. If Minerva-AL accommodates backward blocking, retrieval of  $X$  given  $B$  will be weaker in the backward-blocking condition than in either of the control conditions.

We simulated the procedure in Table 11, 25 independent replications of each condition for each of three levels of the encoding parameter  $L$ . As is shown, Minerva-AL anticipates the backward-blocking effect. Retrieval of  $X$  given  $B$  was weaker in the backward-blocking than in the control conditions. As in the simulation of the forward-blocking effect, the size of the backward-blocking effect increased as a function of  $L$ .

Minerva-AL explains backward blocking as a consequence of discrepancy encoding. In phase 1 of training,  $A$  is established as a retrieval cue for  $B$  (i.e., a within-compound association). Consequently, in phase 2,  $A$  retrieved  $B$  into the echo. Because  $B$  was retrieved into the echo but  $B$  was not presented, the trace stored to memory recorded the absence of  $B$  (i.e.,  $-B$ ) paired with the presence of  $X$  (i.e.,  $+X$ ). This information established the conditions for producing the retrospective revaluation result at test.

When  $B$  was presented at test, the traces that contained its inverse (i.e., the traces that coded  $-B$ ) were activated. Because the similarity of  $B$  to  $-B$  is a negative value, and during retrieval a trace is multiplied by its activation (see Eq. 3), traces with a negative representation of  $B$  and a positive representation of  $X$  were inverted (i.e., a trace that joined  $-B$  and  $+X$  in memory was activated as  $+B$  and  $-X$ ). The sum of the inverted traces produced a negative representation of  $X$  in the echo, thus producing the backward blocking result (i.e.,  $B$  retrieves  $-X$ ).

An important feature of the model's explanation for backward blocking is that the retrieval process produced an inverse representation of  $X$  in the echo even though the memory matrix contained no inverse representations of  $X$  itself. The example underscores the point that in Minerva-AL, a demonstration of learning cannot be identified as an encoding or retrieval effect but, rather, must be considered to be an interaction of the two information-processing stages.

*Recovery from blocking* Another example of retrospective revaluation is recovery from blocking. A recovery from blocking experiment includes three training phases followed by a test. In phase 1 of training,  $A$  is presented, followed by  $X$ . In phase 2 of training,  $AB$  is presented, followed by  $X$ . In phase 3 of training,  $A$  is presented alone. Recovery from blocking is observed when, following all three phases of training,  $B$  behaves as a conditioned exciter of  $X$  (e.g., Blaisdell, Gunther, & Miller, 1999). The result is surprising inasmuch as learning of the  $B \rightarrow X$  relationship is blocked following phase 2 of training but is later expressed

following extinction of the unblocked cue. For our purposes, the recovery from blocking result is important because it demonstrates that an associatively activated cue (i.e.,  $B$  in phase 3 of the procedure) can become a conditioned exciter rather than a conditioned inhibitor of a presented outcome (the result we demonstrated in our simulation of backward blocking).

We applied Minerva-AL to a recovery from blocking procedure. The procedure had three successive training phases followed by a test. In phase 1 of training,  $A$  was presented, followed by  $X$ . In phase 2 of training, the compound cue,  $AB$ , was presented, followed by  $X$ . In phase 3 of training,  $A$  was presented alone. Following training, retrieval of  $X$  given  $B$  was tested. The protocol and control procedures are summarized in Table 12. Recovery from blocking is observed if retrieval of  $X$  given  $B$  is greater than zero in the recovery condition and greater in the recovery condition than in either of the control conditions.

We simulated the three conditions in Table 12, 25 replications for each protocol for each of three levels of the encoding parameter  $L$ . Minerva-AL produced the recovery from blocking effect: Retrieval of  $X$  given  $B$  was greater than zero in the recovery condition and was greater in the recovery condition than in either of the control conditions. The magnitude of the recovery effect correlates with  $L$ .

Our explanation of recovery from blocking follows from the dynamics of storage and retrieval in Minerva-AL. Phase 1 established  $A$  as a retrieval cue of  $X$ , which blocked learning of the  $X|B$  contingency in phase 2. In phase 3,  $A$  retrieved both  $B$  and  $X$  into the echo. Because neither  $B$  nor  $X$  was presented, memory recorded a positive representation of  $A$  paired with negative representations of both  $B$  and  $X$ . At test,  $B$  retrieved the phase 3 traces that coded  $+A$ ,  $-B$ , and  $-X$ . Because traces are multiplied by their activations at retrieval (see Eq. 3), those traces were reinverted at retrieval. The reinverted traces were summed to yield a positive representation of  $X$  in the echo.

*Backward conditioned inhibition* Backward conditioned inhibition (*backward inhibition*) is yet another example of retrospective revaluation. A backward inhibition protocol includes two training phases followed by a test. In phase 1 of training, the learner is presented with pairings of a compound cue,  $AB$ , presented alone. In a subsequent training phase, one element of the compound (i.e.,  $A$ ) is paired with an outcome,  $X$ . Following training, retrieval of  $X$  given  $B$  is tested. Backward inhibition is observed when the learner behaves as though  $B$  signals that  $X$  will not occur (i.e., when  $B$  behaves as a conditioned inhibitor of  $X$ ).

Backward inhibition is important to the study of associative learning for the same reason that backward blocking and recovery from blocking are important: The result demonstrates that the associations of an unrepresented

**Table 12** Simulation of recovery from blocking

Condition	Training			Test	<i>L</i>		
	Phase 1	Phase 2	Phase 3		.33	.67	1.00
Recovery	50 $A \rightarrow X$	50 $AB \rightarrow X$	200 $A$	$X B$	.49 (.03)	.80 (.02)	.99 (.01)
Control (1)	50 $A \rightarrow X$	50 $AB \rightarrow X$	200 $C$	$X B$	.33 (.01)	.37 (.02)	.37 (.01)
Control (2)	50 $A \rightarrow X$	50 $AB \rightarrow X$		$X B$	.33 (.01)	.32 (.01)	.29 (.01)

Note. Means and standard errors computed over 25 independent replications. Standard errors are in parentheses.

but associatively activated cue can be modulated when brought to mind by a presented cue.

We applied Minerva-AL to the backward inhibition procedure. The procedure included two successive training phases followed by a test. In phase 1 of training, compound cue  $AB$  was presented, followed by nothing. In phase 2 of training,  $A$  was presented, followed by  $X$ . Following phase 2, retrieval of  $X$  given  $B$  was tested. The design also included two control conditions (see Table 13). If Minerva-AL accommodates backward inhibition, retrieval of  $X$  given  $B$  should be less than zero in the backward inhibition procedure and should be reliably more negative in the backward inhibition condition than in either of the control procedures.

We simulated the three conditions in Table 13, 25 independent replications for each of three levels of the encoding parameter  $L$ . As is shown, Minerva-AL produced the backward inhibition effect. First, retrieval of  $X$  given  $B$  was less than zero in the backward inhibition condition. Second, retrieval of  $X$  given  $B$  was less in the backward inhibition condition than in either of the control conditions. Consistent with the preceding simulations, the size of the backward-blocking effect was correlated with  $L$ .

Urcelay, Perelmuter, and Miller (2008) evaluated backward inhibition using a summation test. In phase 1 of training,  $AB$  was presented without an outcome. In phase 2,  $A \rightarrow X$  trials were intermixed with  $C \rightarrow X$  trials. At test, retrieval of  $X$  given  $C$ ,  $X$  given  $BC$ , and  $X$  given  $CD$  was tested. They reasoned that if  $B$  had become a conditioned inhibitor,  $X|BC$  ought to be less than both  $X|CD$  and  $X|C$ . The experiment confirmed the predictions. We applied Minerva-AL to Urcelay et al.'s procedure. Minerva-AL made the appropriate prediction (i.e.,  $X|BC < X|CD < X|C$ ).

**Table 13** Simulation of backward conditioned inhibition

Condition	Training		Test	<i>L</i>		
	Phase 1	Phase 2		.33	.67	1.00
Backward inhibition	50 $AB$	50 $A \rightarrow X$	$X B$	-.15 (.03)	-.53 (.06)	-.91 (.01)
Control (1)	50 $AB$		$X B$	.00 (.00)	.00 (.00)	.00 (.00)
Control (2)	50 $AB$	50 $C \rightarrow X$	$X B$	.11 (.01)	.08 (.01)	.05 (.01)

Note. Means and standard errors computed over 25 independent replications. Standard errors are in parentheses.

Minerva-AL's explanation of backward conditioned inhibition is coherent with its explanation of backward blocking and recovery from blocking. In phase 1 of training,  $A$  was established as a retrieval cue for  $B$  (i.e., a within-compound association). Consequently, in phase 2 of training,  $A$  retrieved  $B$ . Because  $B$  was retrieved but  $B$  was not presented, a trace stored to memory included a negative representation of  $B$  (i.e.,  $-B$ ) paired with a positive representation of  $X$ . At test, presenting  $B$  to memory caused those traces to invert (i.e., the  $-B$  and  $+X$  traces were inverted as  $+B$  and  $-X$  traces at retrieval). The activated traces were summed in the echo and yielded an inverse representation of  $X$ . That is,  $B$  retrieved  $-X$ . Thus, as with backward blocking and recovery from blocking, Minerva-AL asserts that backward conditioned inhibition represents an interaction between encoding and retrieval processes.

*Recovery from overexpectation* Recovery from overexpectation is yet another demonstration of retrospective reevaluation that builds on the overexpectation effect. A recovery from overexpectation protocol includes three training phases followed by a test. Training phase 1 involves intermixed presentations of  $A \rightarrow X$  and  $B \rightarrow X$ . Training phase 2 involves presentations of  $AB \rightarrow X$ . Phase 3 of training involves presentations of  $A$  alone (i.e., without  $X$ ). The procedure and its control conditions are summarized in Table 14. Recovery from overexpectation is observed when  $X|B$  is greater in the recovery condition than in the two control conditions.

We conducted 25 replications of the procedure for each of three levels of  $L = .33, .67, \text{ and } 1.0$ . We increased the amount of noise introduced to the echo from the interval  $\pm 0.001$  to the interval  $\pm 0.05$ . As is shown in Table 14,

**Table 14** Simulation of recovery from overexpectation

Condition	Training			Test	<i>L</i>		
	Phase 1	Phase 2	Phase 3		.33	.67	1.00
Recovery	50 $A \rightarrow X / 50 B \rightarrow X$	50 $AB \rightarrow X$	200 $A$	$X B$	.75 (.01)	.78 (.01)	.86 (.01)
Control (1)	50 $A \rightarrow X / 50 B \rightarrow X$	50 $AB \rightarrow X$	200 $C$	$X B$	.73 (.01)	.75 (.01)	.80 (.03)
Control (2)	50 $A \rightarrow X / 50 B \rightarrow X$	50 $AB \rightarrow X$		$X B$	.60 (.01)	.68 (.02)	.80 (.02)

Note. Means and standard errors computed over 25 independent replications. Standard errors are in parentheses.

Minerva-AL anticipates the recovery from the overexpectation effect:  $X|B$  is greater in the recovery than in the control condition. The advantage for  $X|B$  in the recovery condition relative to the first control condition demonstrates that recovery of  $B$  occurs by extinguishing one of its previously partnered cues (i.e.,  $A$ ), but not from extinguishing an unpartnered novel cue (i.e.,  $C$ ). The advantage for  $X|B$  in the recovery condition over the second control condition demonstrates the advantage for  $X|B$  that follows from extinguishing  $A$  in the recovery condition.

Minerva-AL explains recovery from overshadowing in the same way that it explains recovery from blocking. When  $A$  is presented in phase 3 of the procedure, it retrieves a representation of both  $B$  and  $X$  (i.e., because of the within-compound association formed between  $A$  and  $B$  in phase 2 of the procedure). Because  $B$  and  $X$  are both expected but neither is presented, the trace stored to memory records an inverse representation of the two. At test,  $B$  activates those traces, but because  $B$  has a negative similarity to its inverse and because traces are multiplied by their activation at retrieval (see Eq. 3), the  $-B$  and  $-X$  traces are reinverted as  $+B$  and  $+X$  traces at retrieval. Those traces are summed into the echo to yield a more positive representation of  $X$  than otherwise would have been retrieved.

*Recovery from overshadowing* A fifth demonstration of retrospective revaluation is recovery from overshadowing. Overshadowing is observed when retrieval of  $X$  given  $A$  is weaker following  $AB \rightarrow X$  training than following  $A \rightarrow X$  training. Recovery from overshadowing is observed when, subsequently, extinguishing  $B$  improves  $X|A$ . (e.g., Matzel, Schachtman & Miller 1985).

**Table 15** Simulation of recovery from overshadowing

Condition	Training		Test	<i>L</i>		
	Phase 1	Phase 2		.33	.67	1.00
Recovery	50 $AB \rightarrow X$	200 $B$	$X A$	.69 (.02)	.79 (.01)	.93 (.01)
Control	50 $AB \rightarrow X$		$X A$	.61 (.02)	.73 (.02)	.86 (.01)

Note. Means and standard errors computed over 25 independent replications. Standard errors are in parentheses. In the simulations,  $\alpha_A=0.7$ .

We simulated recovery from overshadowing with the procedure in Table 15. In phase 1 of training,  $AB$  was presented, followed by  $X$  ( $\alpha_A = .7$ ). In phase 2 of training, we extinguished  $B$ . At test, we measured retrieval of  $X$  given  $A$ . If recovery from overshadowing occurs,  $A$  will retrieve a better representation of  $X$  in the recovery condition than in the control condition in Table 15.

We conducted 25 replications of each procedure for three values of  $L$ . As is shown in Table 15, Minerva-AL anticipates recovery from overshadowing:  $X|A$  was greater following extinction of  $B$ .

Minerva-AL's explanation of recovery from overshadowing is consistent with its explanations for other examples of retrospective revaluation. In training phase 1, a within-compound association between  $AB$  and  $X$  is established. Thus, in phase 2, presenting  $B$  brings both  $A$  and  $X$  to mind. Because  $A$  and  $X$  are anticipated but neither is presented, the resulting trace records  $-A$  and  $-X$ . At test,  $A$  activates those traces. Because  $A$ 's similarity to its inverse is negative and because traces are multiplied by their activation values in retrieval, the  $-A$  and  $-X$  traces are reinverted at retrieval, and the echo retrieves a positive representation of  $X$ . Once again, the result is not a consequence of encoding or retrieval alone. Rather, recovery from overshadowing emerges from interactions between the two.

*Second-order retrospective revaluation* Denniston, Savastano, and Miller (2001) reported an experiment on retrospective revaluation that included three training phases followed by a test. In phase 1 of training, subjects were exposed to two cues,  $A$  and  $B$ , followed by an outcome,  $X$ . In phase 2 of training, subjects were exposed to two cues,  $B$  and  $C$ , followed by  $X$ . In phase 3,  $A$  was extinguished (i.e.,

presented alone) or it was not presented (i.e., a control condition). The full design is shown in Table 16. Subjects responded less strongly to *C* when *A* was extinguished in phase 3 of the training procedure. The result illustrates second-order retrospective reevaluation. Training with *A* in phase 3 of the procedure influenced responding to *C*. However, *A* and *C* were never paired in the training phase. Rather, *A* had a second-order connection to *C*, through their shared partner, *B*. The result shows that retrospective reevaluation can influence learning of both directly and indirectly previously paired cues. The result places constraints on an account of retrospective reevaluation.

We applied Minerva-AL to Denniston et al.'s (2001) design. As is shown in Table 16, Minerva-AL anticipates second-order retrospective reevaluation:  $X|C$  was smaller after *A* was extinguished.

**Summary** Retrospective reevaluation denotes learning of associatively activated but unrepresented cues. Minerva-AL's explanation of retrospective reevaluation as de novo learning via within-compound associations is consistent with a number of other models (e.g., Dickinson & Burke, 1996; Ghirlanda, 2005; Miller & Matzel, 1988; Van Hamme & Wasserman, 1994; Witnauer & Miller, 2011). However, Minerva-AL offers a unique computational description of how that process unfolds by a combination of discrepancy encoding and trace inversion at retrieval.

## General discussion

Minerva-AL describes the memorial processes involved in representing, storing, and retrieving instances of experience from memory. Each experience is stored in memory as a unique trace (i.e., an instance). Encoding emphasizes unexpected over expected features (i.e., discrepancy encoding). When a probe is presented to memory, traces similar to it are activated, and a sum of the activated traces is retrieved. Because a probe retrieves whole traces, it retrieves the events it has co-occurred with in the past: this is how the model accomplishes associative learning.

Minerva-AL gives a relatively unique description of associative learning (for a related approach, see Blough, 1998). Most theories of associative learning represent the

associations between stimulus units *directly* (i.e., as connection weights). In contrast, Minerva-AL models association as a corollary of retrieval: A probe that is presented to memory retrieves an aggregate of experience that represents its expectations for what will follow. Thus, expectation is an emergent property of retrieval. Critically, the theory does not store associative knowledge, beyond the conjunction of events from individual trials of a learning experiment. Of course, there are limits on this statement. We presume, as does Hintzman (1986), that a learner might formulate and then store an expectation to be retrieved and used later. We note that Thomas et al. (2008) have developed a description of relevant mechanisms for accomplishing deliberate search through memory within the framework of Minerva 2. Adding their hypothesis generation and memory search algorithm to Minerva-AL could model the coordination between feature- and rule-based learning documented in work on associative learning in humans (Shanks & Darby, 1998; Tangen & Allan, 2004).

An instance-based account of learning and memory has several strengths. First, differently from the summative learning theories, an instance-based model acknowledges evidence that nonhuman animals have good long-term memory for the specific (Fagot & Cook, 2006; Vaughan & Greene, 1984). Second, Minerva-AL distinguishes encoding from retrieval, a point that has received rare consideration in formal learning theory (see, however, Bouton, 1993; Miller, 2006; Miller & Matzel, 1988; Stout & Miller, 2007). For example, consider how Minerva-AL explains backward blocking. At test, traces in memory contain representations of positive *X*. Yet, at retrieval, *B* retrieves negative *X* through a process of trace inversion. The example demonstrates that Minerva-AL's knowledge is not in memory but, rather, its expectations emerge from memory during retrieval. Third, examining an instance-based approach to learning and memory speaks to the present controversy over episodic memory in nonhuman animals (Babb & Crystal, 2006; Clayton & Dickinson, 1998). To clarify our contribution to this discussion, our instance-based model is not an account of episodic memory, it is an account of learning that assumes memory for the events of individual learning trials. The distinction is a critical one. By Tulving's (2002) terms, episodic memory

**Table 16** Simulation of second order retrospective reevaluation

Condition	Training			Test	<i>L</i>		
	Phase 1	Phase 2	Phase 3		.33	.67	1.0
Second-order	50 $AB \rightarrow X$	50 $BC \rightarrow X$	50 <i>A</i>	$X C$	.36 (.02)	.08 (.05)	-.21 (.10)
Control	50 $AB \rightarrow X$	50 $BC \rightarrow X$		$X C$	.46 (.02)	.38 (.02)	.36 (.02)

*Note.* Means and standard errors computed over 25 independent replications. Standard errors are in parentheses.

involves autoegetic consciousness—mental reenactment and reexperiencing of episodes in one's personal past. Minerva-AL has no such capability, even though memory consists of "episodes." It remains a point for debate whether instance-based representation can accommodate examples of episodic-like memory in nonhumans. Fourth, many researchers have argued in favor of separate systems for learning and memory—one system that handles associative learning and another that handles memory for the specific (Knowlton, Ramus, & Squire, 1992). Minerva-AL contradicts this position and argues, instead, that learning emerges from parallel retrieval of the specific, a difference that has implications for understanding the structure of memory and the epistemic status of learning (see McClelland & Rumelhart, 1986). Finally, models of associative learning and models of human memory rely on different formalisms. Our reanalysis of associative learning using a model of human memory can serve as a conceptual bridge to help human memory theorists consider the implications of associative learning in their own work and to help learning theorists reconsider their own ideas through the tools of human memory research. Admittedly, there are critical differences in human memory and animal learning, but there is also a great deal of overlap to inform both fields of study (see Bouton & Moody, 2004).

Despite Minerva-AL's many successes, the model commits notable mispredictions. For example, the model incorrectly predicts that after *B* is established as a conditioned inhibitor of *X*, unreinforced training with *B* ought to extinguish the inhibition. Minerva-AL has trouble with peak shift (see McLaren, Bennett, Guttman-Nahir, Kim, & Mackintosh, 1995). Minerva-AL predicts that some examples of retrospective revaluation (e.g., backward blocking) ought to be strong and reliable when, in fact, retrospective revaluation is experimentally elusive (see Blaisdell et al., 1999; Dopson, Pearce, & Haselgrove, 2009; Holland, 1999). Minerva-AL fails to acknowledge mediated conditioning due to its explanation for retrospective revaluation—a confusion recognized elsewhere (see Dwyer, 1999; Graham, Jie, Chan, McLaren, & Wills, 2011). Augmented versions of Minerva-AL can produce second-order conditioning and relative cue validity, but those capabilities require modifications. Minerva-AL does not handle learned inattention and highlighting (Kruschke, 2005; Kruschke & Blair, 2000). Minerva-AL does not distinguish cue-timing results such as the distinction between delay and trace conditioning (see Vogel, Castro, & Saavedra, 2004, for a discussion on the differences between trial-level and real-time models). Finally, even though Minerva 2 and Minerva-AL share principles, the two models differ enough that Minerva 2 does not predict associative learning, and, in some cases, Minerva-AL contradicts the predictions of its parent

theory. We intend to explore all of these problems in future work.

By any rational analysis, learning and memory are indivisible (Ebbinghaus, 1885/1964; Pavlov, 1927). Yet, following the cognitive revolution, theories of learning focused on associative processes, whereas theories of memory focused on information processing. Despite the historical divergence, research in the two traditions has produced provocative consistencies. Humans and other animals exhibit blocking and overshadowing (e.g., Chapman & Robbins, 1990; Dickinson & Shanks, 1985; Kruschke & Blair, 2000), learn complex categories (Herrnstein, Loveland, & Cable, 1976; Mackintosh, 1995; Pearce, 1988), exhibit sensitivity to information content (Froehlich, Herbranson, Loper, Wood & Shimp 2004; Herbranson & Shimp, 2008; Hyman, 1953), show context-dependent learning, and appear to reason about abstract relations among cues (Beckers, Miller, De Houwer, & Urushihara, 2006).

Authorities who have recognized the consistencies have worked to describe learning principles that are general and cut across species. To name a few, Pearce (1988) developed an explanation for prototype learning in pigeons by appealing to a model of human memory. Blough (1998) developed an instance-based model of discrimination in pigeons, using signal detection theory. Gluck and Bower (1988) adapted the Rescorla–Wagner model (Rescorla & Wagner, 1972) to explain category learning in humans. Kruschke (2001) developed connectionist architectures to explain associative learning in humans. Provided a thoughtful review on the associative and cognitive explanations of contingency learning. Allan and her colleagues used the Rescorla–Wagner model to explain enduring color-contingent aftereffects in perception (e.g., Allan & Siegel, 1991; Siegel, Allan, & Eissenberg, 1992).

The work here follows this tradition. The Minerva 2 model was invented to explain human memory. We have shown that a minor variant of the theory also happens to explain associative learning, at least when the process of discrepancy encoding is added. We are comforted that a theory of memory can speak to the problem of learning.

**Author Notes** Randall K. Jamieson, Department of Psychology, University of Manitoba; Matthew J. C. Crump, Department of Psychology, Brooklyn College – CUNY; Samuel D. Hannah, Faculty of Social & Behavioral Sciences, University of Queensland. This research was supported by grants from the Natural Sciences and Engineering Council of Canada to both R.K.J. and M.J.C.C. All simulations were written in Fortran90, compiled using the SUN Microsystems Fortran90 compiler, and run on a Sun Fire X4600 M2 server. The code is available from R.K.J. We thank Lee Brooks, Lorraine Allen, and Shep Siegel. Correspondence concerning this article should be addressed to Randall K. Jamieson, Department of Psychology, University of Manitoba, Winnipeg, MB, Canada, R3T 2N2. Electronic correspondence should be sent to jamiesor@cc.umanitoba.ca.

## References

- Allan, L. G., & Siegel, S. (1991). Characteristics of the indirect McCollough effect. *Perception & Psychophysics*, *50*, 249–257.
- Arndt, J., & Hirshman, E. (1998). True and false recognition in MINERVA2: Explanations from a global matching perspective. *Journal of Memory and Language*, *39*, 371–391.
- Ayres, J. J. B., Moore, J. W., & Vigorito, M. (1984). Hall and Pearce negative transfer: Assessments in conditioned suppression and nictating membrane conditioning experiments. *Animal Learning and Behavior*, *12*, 428–438.
- Babb, S. J., & Crystal, J. D. (2006). Episodic-like memory in the rat. *Current Biology*, *16*, 1317–1321.
- Beckers, T., Miller, R., De Houwer, J., & Urushihara, K. (2006). Reasoning rats: Forward blocking in Pavlovian animal conditioning is sensitive to constraints of causal inference. *Journal of Experimental Psychology: General*, *135*, 92–102.
- Bellingham, W. P., Gillette-Bellingham, K., & Kehoe, E. J. (1985). Summation and configuration in patterning schedules with the rat and rabbit. *Animal Learning and Behavior*, *13*, 152–164.
- Benjamin, A. S. (2010). Representational explanations of “process” dissociations in recognition: The DRYAD theory of aging and memory judgments. *Psychological Review*, *117*, 1055–1079.
- Blaisdell, A. P., Gunther, L. M., & Miller, R. R. (1999). Recovery from blocking achieved by extinguishing the blocking CS. *Animal Learning and Behavior*, *27*, 63–76.
- Blough, D. S. (1998). Context reinforcement degrades discriminative control: A memory approach. *Journal of Experimental Psychology: Animal Behavior Processes*, *24*, 185–199.
- Bouton, M. E. (1993). Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. *Psychological Bulletin*, *114*, 80–99.
- Bouton, M. E., & Moody, E. W. (2004). Memory processes in classical conditioning. *Neuroscience and Biobehavioral Reviews*, *28*, 663–674.
- Brooks, L. R. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 169–211). Hillsdale: Lawrence Erlbaum Associates, Inc.
- Brooks, L. R. (1987). Decentralized control of categorization: The role of prior processing episodes. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 141–174). Cambridge: Cambridge University Press.
- Chapman, G. B., & Robbins, S. J. (1990). Cue interaction in human contingency judgment. *Memory and Cognition*, *18*, 537–545.
- Clark, S. E. (1997). A familiarity-based account of confidence-accuracy inversions in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 232–238.
- Clayton, N. S., & Dickinson, A. (1998). Episodic-like memory during cache recovery by scrub-jays. *Nature*, *395*, 272–274.
- Cole, R. P., Barnet, R. C., & Miller, R. R. (1997). An evaluation of conditioned inhibition as defined by Rescorla’s two-test strategy. *Learning and Motivation*, *28*, 323–341.
- Denniston, J. C., Savastano, H. I., & Miller, R. R. (2001). The extended comparator hypothesis: Learning by contiguity responding by relative strength. In R. R. Mowrer & S. B. Klein (Eds.), *Handbook of contemporary learning theories* (pp. 65–117). Hillsdale: Erlbaum.
- Dickinson, A., & Burke, J. (1996). Within-compound associations mediate the retrospective reevaluation of causality judgements. *Quarterly Journal of Experimental Psychology*, *37*, 397–416.
- Dickinson, A., & Shanks, D. (1985). Animal conditioning and human causality judgment. In L.-G. Nilsson & T. Archer (Eds.), *Perspectives on learning and memory* (pp. 167–191). Hillsdale: Erlbaum.
- Dopson, J. C., Pearce, J. M., & Haselgrove, M. (2009). Failure of retrospective reevaluation to influence blocking. *Journal of Experimental Psychology: Animal Behavior Processes*, *35*, 473–484.
- Dougherty, M. R. P., Gettys, C. F., & Ogden, E. E. (1999). MINERVA-DM: A memory processes model for judgments of likelihood. *Psychological Review*, *106*, 180–209.
- Dwyer, D. M. (1999). Retrospective reevaluation or mediated conditioning? The effect of different reinforcers. *Quarterly Journal of Experimental Psychology*, *52*, 289–306.
- Ebbinghaus, H. (1885/1964). *Memory: A contribution to experimental psychology* (H. A. Ruger, Trans.). New York: Dover. Original work published 1885.
- Fagot, J., & Cook, R. G. (2006). Evidence for large long-term memory capacities in baboons and pigeons and its implications for learning and the evolution of cognition. *Proceedings of the National Academy of Sciences*, *103*, 17564–17567.
- Froehlich, A. L., Herbranson, W. T., Loper, J. D., Wood, D. M., & Shimp, C. P. (2004). Anticipating by pigeons depends on local statistical information in a serial response time task. *Journal of Experimental Psychology: General*, *133*, 31–45.
- Gallistel, C. R., Balsam, P. D., & Fairhurst, S. (2004). The learning curve: Implications of a quantitative analysis. *Proceedings of the National Academy of Sciences*, *101*, 13124–13131.
- Ghirlanda, S. (2005). Retrospective reevaluation as simple associative learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *31*, 107–111.
- Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, *117*, 225–244.
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*, 251–279.
- Graham, S., Jie, H. L., Chan, H.-M., McLaren, I. P. L., & Wills, A. J. (2011). Simultaneous backward conditioned inhibition and mediated conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, *37*, 241–245.
- Hall, G., & Channell, S. (1985). Differential effects of contextual change on latent inhibition and on the habituation of an orienting response. *Journal of Experimental Psychology: Animal Behavioral Processes*, *11*, 470–481.
- 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.
- 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., Gharraei, S., & Westbrook, R. F. (2008). Negative patterning is easier than a biconditional discrimination. *Journal of Experimental Psychology: Animal Behavioral Processes*, *34*, 494–500.
- Herbranson, W. T., & Shimp, C. P. (2008). Artificial grammar learning in pigeons. *Learning & Behavior*, *36*, 116–137.
- Herrnstein, R. J., Loveland, D. H., & Cable, C. (1976). Natural concepts in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, *2*, 285–302.
- Hicks, J. L., & Starns, J. J. (2006). Remembering source evidence from associatively related items: Explanations from a global matching model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 1164–1173.
- Hintzman, D. L. (1986). “Schema abstraction” in a multiple-trace memory model. *Psychological Review*, *93*, 411–428.
- Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, *95*, 528–551.
- Holland, P. C. (1999). Overshadowing and blocking as acquisition deficits: No recovery after extinction of overshadowing or

- blocking cues. *Quarterly Journal of Experimental Psychology*, 52, 307–333.
- Hyman, R. (1953). Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology*, 45, 188–196.
- Jamieson, R. K., Hannah, S. D., & Crump, M. J. C. (2010). A memory based account of retrospective reevaluation. *Canadian Journal of Experimental Psychology*, 64, 153–164.
- Jamieson, R. K., Holmes, S., & Mewhort, D. J. K. (2010). Global similarity predicts dissociation of classification and recognition: Evidence questioning the implicit/explicit learning distinction in amnesia. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 1529–1535.
- Jamieson, R. K., & Mewhort, D. J. K. (2009a). Applying an exemplar model to the artificial-grammar task: Inferring grammaticality from similarity. *Quarterly Journal of Experimental Psychology*, 62, 550–575.
- Jamieson, R. K., & Mewhort, D. J. K. (2009b). Applying an exemplar model to the serial reaction time task: Anticipating from experience. *Quarterly Journal of Experimental Psychology*, 62, 1757–1783.
- Jamieson, R. K., & Mewhort, D. J. K. (2010). Applying an exemplar model to the artificial-grammar task: String-completion and performance for individual items. *Quarterly Journal of Experimental Psychology*, 63, 1014–1039.
- Jamieson, R. K., & Mewhort, D. J. K. (2011). Grammaticality is inferred from global similarity: A reply to Kinder (2010). *Quarterly Journal of Experimental Psychology*, 64, 209–216.
- Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. In B. A. Campbell & R. M. Church (Eds.), *Punishment* (pp. 279–296). New York: Appleton-Century-Crofts.
- Knowlton, B., Ramus, S., & Squire, L. (1992). Intact artificial grammar learning in amnesia: Evidence for parallel learning of item-specific and general information. *Psychological Science*, 3, 172–179.
- Kremer, E. F. (1978). Rescorla–Wagner model: Losses in associative strength in compound conditioned stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, 4, 22–36.
- Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of learning. *Psychological Review*, 99, 22–44.
- Kruschke, J. K. (1996). Base rates in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 3–26.
- Kruschke, J. K. (2001). Toward a unified model of attention in associative learning. *Journal of Mathematical Psychology*, 45, 812–863.
- Kruschke, J. K. (2005). Learning involves attention. In G. Houghton (Ed.), *Connectionist models in cognitive psychology* (pp. 113–140). Hove: Psychology Press.
- Kruschke, J. K., & Blair, N. J. (2000). Blocking and backward blocking involve learned inattention. *Psychonomic Bulletin and Review*, 7, 636–645.
- Kwantes, P. J. (2005). Using context to build semantics. *Psychonomic Bulletin and Review*, 12, 703–710.
- Kwantes, P. J., & Mewhort, D. J. K. (1999). Modeling lexical decision and word naming as a retrieval process. *Canadian Journal of Experimental Psychology*, 53, 306–315.
- Kwantes, P. J., & Neal, A. (2006). Why people underestimate  $y$  when extrapolating in linear functions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 1019–1030.
- Lober, K., & Lachnit, H. (2002). Configural learning in human Pavlovian conditioning: Acquisition of a biconditional discrimination. *Biological Psychology*, 59, 163–168.
- Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, 95, 492–527.
- Logan, G. D. (2002). An instance theory of attention and memory. *Psychological Review*, 109, 376–400.
- Mackintosh, N. J. (1971). An analysis of overshadowing and blocking. *Quarterly Journal of Experimental Psychology*, 23, 118–125.
- Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, 82, 276–298.
- Mackintosh, N. J. (1995). Categorization by people and pigeons: The twenty-second Bartlett memorial lecture. *Quarterly Journal of Experimental Psychology*, 48, 193–214.
- Matzel, L. D., Schachtman, R. R., & Miller, R. R. (1985). Recovery of an overshadowed association achieved by extinction of the overshadowing stimulus. *Learning and Motivation*, 16, 398–412.
- McClelland, J. L., & Rumelhart, D. E. (1986). A distributed model of human learning and memory. In J. L. McClelland, D. E. Rumelhart, & The PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 2: Psychological and biological models* (pp. 170–215). Cambridge: MIT Press.
- McLaren, I. P. L., Bennett, C. H., Guttman-Nahir, T., Kim, K., & Mackintosh, N. J. (1995). Prototype effects and peak shift in categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 662–673.
- Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, 85, 207–238.
- Miller, R. R. (2006). Challenges facing contemporary associative approaches to acquired behavior. *Comparative Cognition & Behavior Reviews*, 1, 77–93.
- Miller, R. R., & Matute, H. (1996). Biological significance in forward and backward blocking: Resolution of a discrepancy between animal conditioning and human causal judgment. *Journal of Experimental Psychology: General*, 125, 370–386.
- Miller, R. R., Barnet, R. C., & Grahame, N. J. (1995). Assessment of the Rescorla–Wagner model. *Psychological Bulletin*, 117, 363–386.
- Miller, R. R., & Matzel, L. D. (1988). The comparator hypothesis: A response rule for the expression of associations. In G.H. Bower (Ed.), *The psychology of learning and motivation, Vol. 22* (pp. 51–92). San Diego, CA: Academic Press.
- Moore, J. W. (1972). Stimulus control: Studies of auditory generalization in rabbits. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current theory and research* (pp. 206–230). New York: Appleton-Century-Crofts.
- 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 and Behavior*, 29, 36–45.
- Nelson, J. B., & Sanjuan, M. C. (2006). A context-specific latent-inhibition effect in a human conditioned-suppression task. *Quarterly Journal of Experimental Psychology*, 59, 1003–1020.
- Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, 115, 39–57.
- Pavlov, I. P. (1927). *Conditioned reflexes*. London: Oxford University Press.
- Pearce, J. M. (1988). Stimulus generalization and the acquisition of categories by pigeons. In L. Weiskrantz (Ed.), *Thought without language* (pp. 132–152). Oxford: Oxford University Press.
- Pearce, J. M. (1994). Similarity and discrimination: A selective review and a connectionist model. *Psychological Review*, 101, 587–607.
- Pearce, J. M. and Hall G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, 87, 532–552.
- Redhead, E. S., & Pearce, J. M. (1995). Similarity and discrimination learning. *Quarterly Journal of Experimental Psychology*, 48B, 46–66.
- Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, 3, 382–407.

- Rescorla, R. A. (1969). Conditioned inhibition of fear resulting from negative CS–US contingencies. *Journal of Comparative and Physiological Psychology*, *67*, 504–509.
- Rescorla, R. A. (1971). Summation and retardation tests of latent inhibition. *Journal of Comparative and Physiological Psychology*, *75*, 77–81.
- Rescorla, R. A. (1973). Evidence for the “unique stimulus” account of configural conditioning. *Journal of Comparative and Physiological Psychology*, *85*, 331–338.
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II* (pp. 64–99). New York: Appleton-Century-Crofts.
- Savastano, H. I., Yin, H., Barnet, R. C., & Miller, R. R. (1998). Temporal coding in Pavlovian conditioning: Hall–Pearce negative transfer. *Quarterly Journal of Experimental Psychology*, *51*, 139–153.
- Shanks, D. R., & Darby, R. J. (1998). Feature- and rule-based generalization in human associative learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *4*, 405–415.
- Siegel, S., Allan, L. G., & Eissenberg, T. (1992). The associative basis of contingent color aftereffects. *Journal of Experimental Psychology: General*, *121*, 79–94.
- Stout, S. C., & Miller, R. R. (2007). Sometimes-competing retrieval (SOCR): A formalization of the comparator hypothesis. *Psychological Review*, *114*, 759–783.
- Tangen, J. M., & Allan, L. G. (2004). Cue interaction and judgements of causality: Contributions of causal and associative processes. *Memory and Cognition*, *32*, 107–124.
- Thomas, R. P., Dougherty, M. R., Sprenger, A. M., & Harbison, J. I. (2008). Diagnostic hypothesis generation and human judgment. *Psychological Review*, *115*, 155–185.
- Tolman, E. C. (1932). *Purposive behavior in animals and men*. New York: Appleton-Century-Crofts.
- Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, *53*, 1–25.
- Urcelay, G. P., Perelmuter, O., & Miller, R. R. (2008). Pavlovian backward conditioned inhibition in humans: Summation and retardation tests. *Behavioral Processes*, *77*, 299–305.
- Van Hamme, L. J., & Wasserman, E. A. (1994). Cue competition in causality judgments: The role of nonpresentation of compound stimulus elements. *Learning and Motivation*, *25*, 127–151.
- Vaughan, W., & Greene, S. L. (1984). Pigeon visual memory capacity. *Journal of Experimental Psychology: Animal Behavior Processes*, *10*, 256–271.
- Vogel, E. H., Castro, M. E., & Saavedra, M. A. (2004). Quantitative models of Pavlovian conditioning. *Brain Research Bulletin*, *63*, 173–202.
- von Restorff, H. (1933). Über die Wirkung von Bereichsbildungen im Sperunfeld [On the effect of spheres formations in the trace field]. *Psychologische Forschung*, *18*, 448–456.
- Voss, J. L. (2009). Long-term associative memory capacity in man. *Psychonomic Bulletin and Review*, *16*, 1076–1081.
- Wagner, A. R. (1981). SOP: A model of automatic memory processing in animal behavior. In N. E. Spear & R. R. Miller (Eds.), *Information processing in animals: Memory mechanisms* (pp. 5–47). Hillsdale: Erlbaum.
- Wasserman, E. A., & Miller, R. R. (1997). What's elementary about associative learning? *Annual Review of Psychology*, *48*, 573–607.
- Whittlesea, B. W. A., & Williams, L. D. (2000). The source of feelings of familiarity: The discrepancy-attribution hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 547–565.
- Whittlesea, B. W. A., & Williams, L. D. (2001a). The discrepancy attribution hypothesis: I. The heuristic basis of feelings and familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 3–13.
- Whittlesea, B. W. A., & Williams, L. D. (2001b). The discrepancy attribution hypothesis: II. Expectation, uncertainty, surprise, and feelings of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 14–33.
- Williams, B. A., & McDevitt, M. A. (2002). Inhibition and superconditioning. *Psychological Science*, *13*, 454–459.
- Witnauer, J. E., & Miller, R. R. (2011). The role of within-compound associations in learning about absent cues. *Learning & Behavior*, *39*, 146–162.
- Young, M. E., Wasserman, E. A., Johnson, J. L., & Jones, F. L. (2000). Positive and negative patterning in human causal learning. *Quarterly Journal of Experimental Psychology*, *53*, 121–138.