Canadian Journal of Experimental Psychology
2010, Vol. 64, No. 3, 153–164

1196-1961/10/$12.00 DOI: 10.1037/a0020886

A Memory-Based Account of Retrospective Revaluation

Randall K. Jamieson
University of Manitoba

Samuel D. Hannah
University of Queensland

Matthew J. C. Crump
Vanderbilt University

We adapt an instance model of human memory, Minerva 2, to simulate retrospective revaluation. In the account, memory preserves the events of individual trials in separate traces. A probe presented to memory contacts all traces in parallel and causes each to become active. The information retrieved from memory is the sum of the activated traces. Learning is modelled as a process of cued-recall; encoding is modelled as a process of differential encoding of unexpected features in the probe (i.e., expectancy-encoding). The model captures three examples of retrospective revaluation: backward blocking, recovery from blocking, and backward conditioned inhibition. The work integrates an understanding of human memory and complex associative learning.

Keywords: Instance theory, associative learning, retrospective revaluation, Minerva 2

In retrospective revaluation, unpresented but associatively activated cues are associated. One example of retrospective revaluation is backward blocking. A simple demonstration of backward blocking involves two successive training phases followed by a test. In phase one of training, a cue compound, *AB*, is presented followed by an outcome, *X*. In phase two, *A* is presented followed by *X*. Backward blocking is observed if, following training, *B* is a weak exciter of *X* (Shanks, 1985). Of course, the result contradicts common sense. In the training phases, *B* reliably predicts *X*. Yet, subjects behave as if the opposite was true (i.e., that *B* does not predict *X*). Retrospective revaluation implicates a role of memory in learning. Van Hamme and Wasserman (1994; see also Dickinson & Burke, 1996) argued that in phase one of the backward blocking procedure a within-compound association forms between *A* and *B*. In phase two, the within-compound association causes *A* to retrieve *B*. Because *B* is retrieved, it can develop an inhibitory link to *X*. Melchers, Lachnit, and Shanks (2004) proposed a different albeit related memory-based explanation. In their account, the presentation of *A* in phase two elicits covert rehearsal of phase one trials, and backward blocking falls out of the covert rehearsal process. Here, we develop a novel explanation of retrospective revaluation using an instance model of human memory (Hintzman, 1984, 1986, 1988).

Instance theories of learning and memory operate from a premise that the individual experience (i.e., the instance) is the primitive unit of knowledge and that learning represents the accumulation and deployment of instances from memory. Brooks (1978, 1987) was amongst the first to champion the view. Medin and Schaffer (1978) were amongst the first to formalize it. Hintzman's (1984, 1986, 1988) Minerva 2 model and Nosofsky's (1986) Generalized Context Model represent formal first-generation accounts of the instance-based view of memory. Kruschke's (1992, 1996, 2001) ALCOVE, ADIT, and EXIT models and Logan's (1988, 2002) ITAM model are modern extensions of the exemplar-based view of learning. Whereas different instance-based theories differ in their details, all agree that the instance is the fundamental unit of knowledge and that a competent theory of learning must include an account of how instances are stored and retrieved from memory.

In this paper, we adapt Hintzman's (1986, 1988) Minerva 2 instance-based model of human memory to an analysis of retrospective revaluation. In short, we propose that traces of individual trials are stored in memory, that learning is driven by a process of expectancy-encoding, that decisions about associative strength fall out of a process of parallel cued-recall, and that retrospective revaluation follows from a process of trace-inversion at retrieval.

Minerva 2

Minerva 2 is a classic instance-based theory of human memory. The theory was developed to understand episodic-recognition and frequency-judgement (Hintzman, 1984, 1986, 1988). Minerva 2 has since been applied to a wide range of phenomena from the study of human memory (Arndt & Hirschman, 1998; Clark, 1997; Dougherty, Gettys, & Ogden, 1999; Goldinger, 1998; Hintzman, 1987; Jamieson & Mewhort, 2009a, 2009b, 2010; Jamieson, Holmes, & Mewhort, in press; Kwantes, 2005; Kwantes & Mewhort, 1999; Kwantes & Neal, 2006).

Editor's Note. The paper was part of the Past-President's Symposium held at the joint meeting of CSBBCS and EPS in York, UK, July, 2009.—DJKM

Randall K. Jamieson, Department of Psychology, University of Manitoba; Samuel D. Hannah, School of Psychology, University of Queensland; Matthew J. C. Crump, Department of Psychological Sciences, Vanderbilt University.

We thank Lee Brooks, Lorraine Allen, and Shep Siegel. The research was supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada to Randall K. Jamieson.

Correspondence concerning this article should be addressed to Randall K. Jamieson, Department of Psychology, University of Manitoba, Winnipeg, MB, Canada, R3T 2N2. E-mail: randy_jamieson@umanitoba.ca

Informally, Minerva 2 is a theoretical framework that articulates the memorial processes involved in representing, storing, and retrieving instances of experience. A first central assumption of the model is that each individual experience is represented in memory by a unique trace. A second central aspect of Minerva 2 is the retrieval process. In the model, retrieval is cue-driven and parallel. When a cue (i.e., a memory probe) is presented, it activates all traces in memory. Each trace's activation is in proportion to its similarity to the probe. The information retrieved from memory is the sum of the activated instances, a structure called *the echo*. Because the probe retrieves traces similar to it, a probe will retrieve a representation of itself from memory. Because a probe retrieves whole traces, a probe also retrieves events it has co-occurred with in the past. This is the mechanism that Minerva 2 uses to accomplish cued recall, and it is the mechanism that we will use 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 specific stimulus properties (e.g., has wings) or can be read as information states (e.g., neural potentials). Each feature takes one of three discrete values: $+1$, -1, or 0. A value of $+1$ or -1 indicates the feature is relevant to the stimulus description; a value of 0 indicates the feature is either indeterminate or irrelevant to the stimulus description.

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

Memory, M, is a two dimensional matrix. Each row in the matrix stores an instance. Each column corresponds to a feature. Encoding an event vector, E, involves copying the elements of E to a row in the memory matrix. The model accommodates variation in the quality of encoding by varying the probability with which each element in a stimulus vector is copied to memory. The probability of storing each element in memory is a model parameter, L. An element that is not stored is copied to memory as a value of 0. Thus, as L increases, items are stored more completely in memory.

In the model, all retrieval is cued. When a cue is presented, it activates each memory trace in proportion to its similarity to the cue. In Minerva 2, similarity of the probe, P, to trace i in memory, M_i, is computed as,

$$S_i = \frac{\sum_{j=1}^{n} P_j \times M_{ij}}{n_R}, \tag{1}$$

where P_j is the value of the *jth* feature in the probe, M_{ij} is the value of the *jth* feature of the *ith* 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, is -1 when the row is opposite to the probe, and is 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,

$$A_i = S_i^3. \tag{2}$$

In principle, the probe activates all traces in memory. However, the nonlinear activation function ensures that traces very similar to the probe are activated much more strongly than traces that are moderately similar or that are dissimilar to the probe.

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

$$C_j = \sum_{i=1}^{m} A_i \times M_{ij}. \tag{3}$$

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

Quality of cued-recall is indexed by, first, normalizing values in the echo,

$$C_j' = \frac{C_j}{\max|C_{1..n}|}, \tag{4}$$

and, then, computing the similarity between the normalized echo and the target associate, X:

$$X|P = \frac{\sum_{j=1}^{n} X_j \times C_j'}{n_R}, \tag{5}$$

where X is a target associate, P is the probe, and n_R is the number of nonzero features in X. 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. The value $X|P$ behaves like a Pearson correlation coefficient. If the probe retrieves X perfectly, $X|P = 1$. If the probe does not retrieve X, $X|P = 0$. If the probe retrieves a perfect inverse (i.e., opposing) representation of X, $X|P = -1$.

Now that we have described the Minerva 2 model, we move to a description of how we adapted the model to the problem of associative learning.

Minerva-AL

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 regard for the order in which list-items were presented or potential encoding dependencies amongst list-items. The same is true in studies of categorization and cued-recall.

Whereas independent encoding of list-items allows simulation of performance in memory experiments, it is insufficient to simulate learning: In a learning experiment, the problem of interest is how memory of events from preceding trials influences processing of and memory for events on a present trial. Therefore, we address learning by adapting the Minerva 2 model so that memory of a present trial is influenced by memory of preceding trials. We will call the adapted model, Minerva-AL.

The key difference between Minerva-AL and Minerva 2 is in how Minerva-AL encodes an experience. In Hintzman's (1984, 1986) original Minerva 2 model, memory for a trial is established by copying the event vector to a row in the memory matrix. In Minerva-AL, memory for a trial is determined as the difference between the event vector and the echo retrieved. By encoding differences between the event vector and the echo, memory of preceding trials (i.e., represented in the echo) has influence on what is learnt on the trial (i.e., represented in the event vector). In short, unexpected information (i.e., information in the event vector that is not retrieved in the echo) is encoded more strongly than expected information (i.e., information in the event vector that is retrieved in the echo). Because the encoding operation in Minerva-AL is driven by a concept of expectancy, we call the operation *expectancy-encoding*.

In Minerva-AL, expectancy-encoding is implemented using subtraction,

$$M_{ij} = E_j - C_j', \tag{6}$$

where i indexes the row in memory, j indexes the features of the vector representations, M is the memory matrix, E is the event vector, and C' is the echo. We retain Minerva 2's probabilistic encoding rule: $M_{ij} = E_j - C_j'$ with probability L and $M_{ij} = 0$ with probability $1 - L$.

To illustrate expectancy-encoding, imagine a learning trial where A is presented followed by X. In the example, A and X are represented by four features so that $A = [1, 1, 0, 0]$ and $X = [0, 0, 1, 1]$. Because the trial presents A followed by X, the event vector, E, is equal to $E = A + X = [1, 1, 1, 1]$. Suppose that on trial i in the experiment, 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 show that the model retrieves a strong expectation for X. According to expectancy-encoding (see Equation 6), the information stored to row i in memory will equal $M_i = E - C' = [0.6, 0.9, 0.4, 0.0]$. Note that the most anticipated feature in the echo (feature 4) is encoded as a zero, the second most anticipated feature in the echo (feature 3) is encoded as the second smallest absolute value, and so on.

An important corollary of expectancy-encoding is that Minerva-AL appreciates and encodes violations of its expectations. To illustrate, consider a variation on the example from the preceding paragraph. On trial i, A retrieves $C' = [0.4, 0.1, 0.6, 1.0]$, just as before. However, X is not presented. Thus, in contrast to the preceding example, $E = A = [1, 1, 0, 0]$. In this scenario, the information stored to row i in memory is equal to $M_i = E - C' = [0.6, 0.9, -0.6, -1.0]$. Note that, now, the information encoded to the third and fourth elements of row i in memory take the opposite sign of the original representation for $X = [0, 0, 1, 1]$. This inverse representation of X records the fact that the model expected X but that X did not occur. This property of the model will be pivotal for our eventual explanation of retrospective revaluation.

Of course, we are not the first to argue for the importance of expectancy-encoding. Kamin (1969) and von Restorff (1933) identified surprise as a key principle of learning and memory. Whittlesea and Williams (2000, 2001a, 2001b, 2001b) used violation of expectancy to explain memory-based inference. Rescorla and Wagner (1972) used surprise to model learning in cue competition.

The expectancy-encoding operation required two additional changes to the Minerva 2 model. First, the similarity rule in Minerva 2 (see Equation 1) is tailored to the situation where features of stimulus and memory representations can take one of only three discrete values (+1, 0, −1), but expectancy-encoding allows feature values to vary continuously, between −2 and + 2. We resolve the problem using Kwantes' (2005) solution. He computed similarity between a probe and memory trace using the cosine measure of similarity:

$$S_i = \frac{\sum_{j=1}^{n} P_j \times M_{ij}}{\sqrt{\sum_{j=1}^{n} P_j^2}\sqrt{\sum_{j=1}^{n} M_{ij}^2}}, \tag{7}$$

where P_j is the value of the *jth* feature in the probe, M_{ij} is the value of *jth* feature of the *ith* row in memory, and n is the number of features in the vectors under comparison. The cosine measure of similarity is consistent with the similarity measure used in the Minerva 2 model. However, it normalizes over vector length and, thus, handles the extended range of values in memory traces that follow from the expectancy-encoding rule.

A final change to the model involved adding a randomly sampled value from the interval [−0.001, +0.001] to each element in the echo. The change was pragmatic. Minerva 2 is a model of human memory and so it was designed for single-trial learning. If noise is not added to the echo, Minerva-AL learns too quickly (often, in a single trial).

In the simulations that follow, we simulate associative learning as an example of cued-recall: presenting a cue retrieves an echo, and the echo is assessed for the target outcome. To ease exposition, we denote a cue's ability to retrieve an outcome as a conditional. For example, $X|B$ refers to "retrieval of X given B" and $X|AB$ refers to "retrieval of X given AB". Positive growth in conditional retrieval corresponds to a growing excitatory association between the cue and outcome; negative growth corresponds to a growing inhibitory association between the cue and outcome.

In all of the simulations that follow, we use Hintzman's (1986) scheme for stimulus representation. Events of a trial are coded in an event vector composed of five successive 20-element subfields (i.e., an event vector has 100 elements in total). The first, second, third, and fourth subfields correspond to cues A, B, C, and D, respectively. The fifth subfield corresponds to the target outcome X. We represent cues and outcomes by assigning values of +1 and −1 with equal probability to each of the 20 relevant elements in the stimulus representation. Thus, A is represented by assigning a value +1 or −1 to each of the first 20 elements of a 100-element vector; all remaining elements take a value zero. Stimulus B is represented by assigning a value +1 or −1 to elements 21 through 40 of a 100-element vector, with all other values zero. Stimulus C

is represented by assigning a value +1 or −1 to elements 41 through 60 of a 100-element vector, with all other values zero. Stimulus *D* is represented by assigning a value +1 or −1 to each of the elements 61 through 80 of a 100-element vector, with all other values 0. Outcome *X* is represented by assigning a value +1 or −1 to each of the elements 81 through 100 of a 100-element vector, with all other values 0.[1]

Despite the computational differences between Minerva-AL and Minerva 2—with the critical distinction of expectancy-encoding—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). Finally, information retrieved by a cue is quantified from the echo.

Simple Associative Learning

We first apply Minerva-AL to five elementary learning protocols: acquisition, extinction, backward conditioning, blocking, and conditioned inhibition. We use these simulations to illustrate the model and to show it handles basic associative learning.

Acquisition/extinction. In a simple associative learning procedure, a cue, *A*, is presented followed by an outcome, *X*. After several pairings, the cue elicits anticipation of the outcome: a result called *acquisition*. If the cue is then presented alone (i.e., without the outcome), its ability to elicit anticipation of the outcome fades: a result called *extinction*.

We applied Minerva-AL to an acquisition/extinction protocol that included 200 trials. Trials 1 through 100 were acquisition trials (i.e., *A* presented followed by *X*); Trials 101 through 200 were extinction trials (i.e., *A* presented alone).

At the outset of each trial, memory was probed with *A* and an echo was retrieved. Learning was recorded as retrieval of *X* given *A*. The trial was completed by storing a trace in memory. For acquisition trials, the event vector, *E*, was equal to $A + X$; for extinction trials, $E = A$.

We conducted 25 independent simulations of the procedure for each of three levels of *L* (we varied *L* to illustrate that learning in the theory is modulated-by but is not dependent upon particular values of the parameter). Figure 1 shows retrieval of *X* given *A* over the 200 trials of the protocol. The curves in Figure 1 are averaged over 25 independent replications of the protocol. As shown, Minerva-AL produced negatively accelerated acquisition and extinction curves with the rate of learning systematically correlated with *L*. The curves match the characteristic shapes of averaged learning curves.

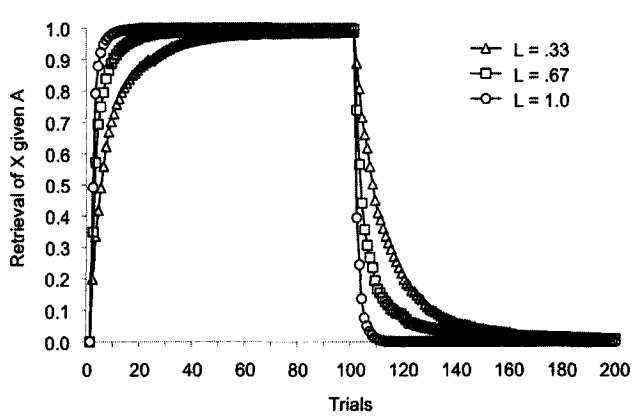

Figure 1. Acquisition (Trials 1 – 100) and extinction (Trials 101 – 200). Means are computed from 25 independent replications of the procedure.

To gain a better understanding of how the model learns, we inspected the trial-to-trial mechanics of the simulation. At the outset of a simulation, memory was empty (a matrix of 0s). Consequently, on Trial 1, *A* retrieved only noise into the echo. Because the echo contained only noise, retrieval of *X* given *A* was approximately zero ($M = 0.003$, $SE = .0167$) and elements in the event vector (i.e., $E = A + X$) were encoded strongly to memory (see Equation 6). On Trial 2, *A* retrieved a noisy version of the trace stored on the preceding trial. Because the trace included information about the pairing of *A* and *X* on Trial 1, retrieval of *X* given *A* improved on Trial 2. On Trial 3, *A* retrieved the traces from both Trials 1 and 2. Because the two traces were summed into the echo, noise in the echo was reduced and a more complete representation of *X* was retrieved. Because each additional trace was summed into the echo, there was a systematic and cumulative benefit over successive trials.

On Trial 101, cue *A* retrieved a near-perfect representation of outcome *X*. However, *X* was not presented. Consequently, the trace that was stored to memory included a representation of *A* paired with an inverse representation of *X* (i.e., -*X*). On Trial 102, cue *A* retrieved the traces from Trials 1 through 101. Because the inverse representation of *X* from Trial 101 was retrieved in the echo, retrieval of *X* given *A* suffered; note the drop in $X|A$ on Trial 102. This process cumulated over the remaining extinction trials and produced a corresponding cumulative impairment of $X|A$. Eventually, retrieval of *X* given *A* was statistically equal to zero indicating that *A* had ceased to elicit anticipation of *X* altogether.

We conducted additional simulations of acquisition and extinction. In one series of simulations, we varied the number of trials in the two phases of the design. Results of these simulations were consistent with the results of the simulations reported in Figure 1: $X|A$ approached 1 in acquisition and 0 in extinction. In another series of simulations, we varied the cue-outcome contingency (i.e., the probability of *X* following *A*). As in the simulations from Figure 1, Minerva-AL predicted asymptotic learning of $X|A$. However, the asymptotic value of $X|A$ closely approximated the probability of *X* given *A* in the simulated protocol. In yet another series of simulations, we varied the amount of noise added to the echo: For example, rather than add a value from the range [−0.001, +0.001] to each element in the echo, we added a value from the range [−0.1, +0.1]. Broadening the range of noise slowed learning; shrinking the range did the opposite.

The results of the simulation in Figure 1 show that Minerva-AL accommodates acquisition and extinction. We now turn to a demonstration that the model handles the distinction between forward and backward conditioning.

Backward conditioning. In a backward conditioning procedure, an outcome, *X*, is presented ahead of a cue, *A*. At test, the

[1] The results of the simulations in this article do not depend on using the +1/−1/0 feature values as in Hintzman's (1986, 1988) scheme. Simulations using other schemes (e.g., sampling values from a Gaussian) give similar results.

learner treats *A* as a weak or inhibitory predictor of *X*. The result is rational: in training, *A* signals that the outcome is ended. The backward conditioning procedure is important because it illustrates that association is asymmetric. Following training with *A* followed by *X*, *A* is a better predictor of *X* than *X* is of *A*. If Minerva-AL is to serve as a model of learning, it should handle the asymmetry.

We applied Minerva-AL to both a forward and a backward conditioning protocol. Both protocols included a training phase followed by a test. In the forward training protocol, *A* was presented followed by *X*. In the backward training protocol, *X* was presented followed by *A*. Following the training phase, we measured retrieval of *X* given *A* (i.e., *X*|*A*). If Minerva-AL distinguishes forward from backward conditioning, *A* ought to be a better retrieval cue for *X* following the forward than following the backward conditioning procedure.

We conducted simulations of the forward and backward conditioning procedures, 25 replications of each procedure for each of three levels of *L*. Following training with the forward conditioning procedure, *X*|*A* was equal to .49 (*SE* = .03), .90 (*SE* = .02), and 0.99 (*SE* = .00), for *L* = .33, .67, and 1.0, respectively. By contrast, following training with the backward training procedure, *X*|*A* was equal to .21 (*SE* = .01), .34 (*SE* = .03), and .73 (*SE* = .03), for *L* = .33, .67, and 1.0, respectively. The simulations conform to the expected difference: *A* is a better retrieval cue for *X* in the forward than in the backward conditioning procedure.

The simulations demonstrate that Minerva-AL distinguishes forward from backward conditioning: a prediction of Minerva-AL that is at odds with Minerva 2. In Minerva 2, encoding is independent of retrieval. Because of that independence, Minerva 2 predicts that *X*|*A* will be equal in the forward and backward conditioning procedures.

Our simulations show that the Minerva-AL model accommodates simple associative learning. We now turn to a more sophisticated learning problem: blocking.

Blocking. Blocking illustrates a process of cue-competition in learning. A classical blocking procedure involves two successive training phases followed by a test (Kamin, 1969; Rescorla & Wagner, 1972). In phase one of training, a cue, *A*, is presented followed by an outcome, *X*. In phase two of training, a cue compound, *AB*, is presented followed by *X*; critically, the cue compound presented in phase two, *AB*, includes the cue presented in phase one of the procedure (in this example, *A*). Following phase two, retrieval of *X* given *B* is tested. A schematic of the procedure is provided in the top row of Table 1. The second and third rows of Table 1 describe relevant control procedures. Blocking is demonstrated if retrieval of *X* given *B* is weaker in the blocking condition than in the control conditions. Minerva-AL must accommodate blocking to stand as competent account of learning.

We simulated the blocking and control procedures described in Table 1, 25 independent replications of each condition and for each of three levels of the encoding parameter *L*. As shown in Table 1, Minerva-AL anticipates the blocking effect (i.e., *X*|*B* was smaller in the blocking condition than in the control conditions). The magnitude of blocking covaries with *L*.

Minerva-AL's explanation for blocking is straightforward. In phase one of training, *A* is established as a retrieval cue for *X*. In phase two, the compound cue *AB* retrieved *X* into the echo. Because *AB* retrieved *X* into the echo, the biggest discrepancy between the echo and the event vector was the presence of *B*. The trace stored to memory, therefore, included a strong representation of *B* but a weak representation of *X*. Consequently, at test, *B* retrieved a weak representation of *X*.

Thus far, we have showed Minerva-AL handles acquisition, extinction, backward conditioning, and blocking. Next, we test Minerva-AL against the problem of conditioned inhibition.

Conditioned inhibition. Conditioned inhibition demonstrates an organism can learn that an outcome will not be presented. A typical conditioned inhibition procedure involves a training phase followed by a test. The training phase includes two types of trials. For half of the training trials, a cue *A* is presented followed by an outcome *X*. For the other half of the training trials, a cue compound *AB* is presented without *X*. The two types of trials are intermixed. At test, retrieval of *X* given *B* is tested. Conditioned inhibition is observed if, following training, *B* behaves as a conditioned inhibitor of *X* whereas *A* behaves as a conditioned exciter of *X*.

We simulated the conditioned inhibition procedure, 25 independent replications for each of three values of *L* (i.e., *L* = .33, .67, and 1.0). Following the training procedure, we measured both retrieval of *X* given *B* and retrieval of *X* given *A*. For *L* = .33, .67, and 1.0, retrieval of *X* given *B* was equal to −.33 (*SE* = .02), −.49 (*SE* = .03), and −.73 (*SE* = .04), respectively. For *L* = .33, .67, and 1.0, retrieval of *X* given *A* was equal to .93 (*SE* = .02), .96 (*SE* = .02), and .97 (*SE* = .01), respectively. Because retrieval of *X* given *B* was reliably less than zero, we conclude that Minerva-AL handles the problem of conditioned inhibition.

In other laboratory demonstrations of conditioned inhibition, researchers use a summation test. This involves two training phases. In phase one of training, the learner is presented with the intermixed *A* → *X* and *AB* → *nothing* trials. In phase two of training a novel cue, *C*, is presented followed by the outcome, *X*. At test, *X*|*BC*, *X*|*C*, and *X*|*CD* are assessed (i.e., *D* is a novel cue not presented in training). If *B* is a conditioned inhibitor of *X*, then *X*|*BC* ought to be less than both *X*|*C* and *X*|*CD* (e.g., Rescorla,

Table 1
Simulation of Blocking: Retrieval of X Given B as a Function of L (Standard Errors in Parentheses)

	Training			Learning rate (*L*)		
Condition	Phase 1	Phase 2	Test	0.33	0.67	1.00
Blocking	50 A → X	50 AB → X	X|B	.24 (.02)	.22 (.02)	.18 (.02)
Control (1)		50 AB → X	X|B	.51 (.02)	.69 (.02)	.84 (.02)
Control (2)	50 C → X	50 AB → X	X|B	.54 (.02)	.69 (.02)	.85 (.02)

Note. Means and standard errors are computed from 25 independent replications of the procedure. Numbers next to cues denote number of trials.

1969, 1971). We tested Minerva-AL in the summation test. The theory made the appropriate prediction. Retrieval of $X|BC$ was less than retrieval of $X|CD$ and retrieval of $X|CD$ was less than retrieval of $X|C$. The differences were present for all values of L.

To understand why Minerva-AL predicts conditioned inhibition of B, we inspected the trial-to-trial dependencies of the simulation. For trials in which A was presented followed by X, traces recorded a representation of A's co-occurrence with X. Because A was part of the probe on AB trials, AB retrieved X. Because X was expected but not presented, a trace was added to memory that recorded $+A$ and $+B$ paired with $-X$. At test, B retrieved those traces and, consequently, included a negative representation of X in the echo.

The simulations reported, thus far, are helpful in that they afford a clear description of the model's mechanics. However, none of the learning procedures we have simulated challenge existing models. To better challenge Minerva-AL, we now apply it to three examples of retrospective revaluation: backward blocking, recovery from blocking, and backward conditioned inhibition. We will show that Minerva-AL handles retrospective revaluation.

Complex Learning: 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 the assertion. One demonstration of retrospective revaluation is backward blocking.

The backward blocking procedure includes two successive training phases followed by a test. In phase one of training, a compound cue AB is paired with an outcome X. In phase two of training, A is paired with X. After phase two of training, retrieval of X given B is tested. Backward blocking is observed if following phase two of training retrieval of X given B is worse in the backward blocking condition than in the control conditions. The top row in Table 2 outlines the backward blocking procedure. Rows 2 and 3 in Table 2 represent relevant control procedures.

At first blush, backward blocking would appear to imply inferential reasoning: "I learnt in phase one that the combination of A and B predicts X. But, in phase two, I learnt that A alone predicts X. Given the two contingencies, I will infer that, despite initial appearances, B must not have been a predictor of X in phase one after all." Despite such an obvious solution to explaining the result with humans, backward blocking has been observed in the behaviour of both honeybees (Blaser, Couvillon, & Bitterman, 2004) and rats (Miller & Matute, 1996). Although it might be exciting to make the leap and declare honeybees and rats capable of inferential reasoning (e.g., Beckers, Miller, De Houwer, & Urushihara, 2006), it is first appropriate to search for a more basic learning process that produces the backward blocking result.

To explain backward blocking without inferential reasoning, Van Hamme and Wasserman (1994) adapted the Rescorla-Wagner model. According to their adapted model, phase one of training establishes A as a retrieval cue for B—a within-compound association. Because of the within-compound association between A and B, A retrieves B in phase two of the training procedure. Because B is retrieved but not presented, it loses associative strength to X acquired in phase one. Van Hamme and Wasserman argue that backward blocking is this loss of associative strength. Because the account does not rely on a process of reasoning, it finesses the problem of attributing complex reasoning to species in which such capabilities are suspect (e.g., honeybees and rats). Nevertheless, the account makes three implicit assumptions. First, it assumes that the learner recognises which cues are absent. Second, it assumes that the learner distinguishes absent from presented cues. Third, it assumes that the learner applies different parameters to the absent and presented cues when updating their respective associative strengths to the outcome.

Whereas we agree with the thrust of Van Hamme and Wasserman's (1994) explanation of backward blocking, their computational solution is problematic. To simulate backward blocking using the modified Rescorla-Wagner theory, the absent and presented cues must be specified for the model and different learning parameters applied to the two kinds. A more ideal computational solution would identify the absent and presented cues and, using that discrimination, accommodate backward blocking. In the next simulation, we test whether Minerva-AL meets the challenge.

Backward blocking. We simulated a standard backward blocking procedure. The procedure had two successive training phases followed by a test. In phase one of training, compound cue AB was presented followed by outcome X. In phase two of training, cue A was presented followed by X. Following phase two, retrieval of X given B was tested. The design included two control conditions; the full design is described schematically in Table 2. If Minerva-AL accommodates backward blocking, retrieval of X given B will be reliably weaker in the backward blocking condition than in the control conditions.

We simulated the procedure, 25 independent replications of each condition for each of three levels of the encoding parameter L. Results of the simulations are presented in Table 2. As shown, Minerva-AL produced the backward blocking effect: Retrieval of X given B was smaller in the backward blocking than in the control

Table 2
Simulation of Backward Blocking: Retrieval of X Given B as a Function of L (Standard Errors in Parentheses)

	Training			Learning rate		
Condition	Phase 1	Phase 2	Test	0.33	0.67	1.00
Backward blocking	50 $AB \rightarrow X$	50 $A \rightarrow X$	$X\|B$	.41 (.03)	.15 (.04)	.04 (.03)
Control (1)	50 $AB \rightarrow X$		$X\|B$	.64 (.01)	.76 (.01)	.88 (.01)
Control (2)	50 $AB \rightarrow X$	50 $C \rightarrow X$	$X\|B$	.66 (.02)	.80 (.01)	.88 (.01)

Note. Means and standard errors are computed from 25 independent replications of the procedure. Numbers next to cues denote number of trials.

conditions. As in the simulation of the classical blocking effect, the size of the backward blocking effect increased as a function of *L*.

To understand how Minerva-AL solved backward blocking, we inspected the mechanics of the simulation. In phase one of training, *A* was established as a retrieval cue for *B* (i.e., a within-compound association). Consequently, in phase two, *A* retrieved *B* into the echo. Because *B* was retrieved into the echo but *B* was not presented, the resulting traces recorded the absence of *B* (i.e., an inverse representation of *B*) paired with the presence of *X* (i.e., a positive representation of *X*). When *B* was presented at test, it activated the traces that contained its inverse. But, because the similarity of *B* to its inverse is negative and at retrieval each trace in memory is multiplied by its activation, traces with a negative representation of *B* and a positive representation of *X* were inverted (i.e., a trace that joined $-B$ and $+X$ was activated as $+B$ and $-X$ during retrieval). Summing the inverted traces produced a negative representation of *X* in the echo and, thereby, produced the backward blocking result (i.e., *B* retrieved $-X$). It is interesting to note that the mechanics involved in retrieval produced an inverse representation of *X* even though the memory matrix itself contained no inverse representations of *X*. Minerva-AL suggests that backward blocking is not an encoding or retrieval effect, but rather an interaction of the two processes.

Minerva-AL's explanation of the backward blocking effect (the formation and exploitation of a within-compound association between *A* and *B*) is consistent up to a point with the one given by Van Hamme and Wasserman (1994). However, in contrast to Van Hamme and Wasserman's model, Minerva-AL describes (a) how *A* is established as a retrieval cue for *B* in phase one of training, (b) how *A* consequently retrieves *B* in phase two of training, (c) how a violation of the model's expectation of *B* in phase two is represented and encoded to memory, and (d) how the traces in memory interact in the process of retrieval to produce the backward blocking result. The simulation clarifies how learning about the *A*-*X* contingency in phase two of the procedure causes indirect learning about the *B*-*X* contingency, even though *B* is not presented.

Recovery from blocking. A second example of retrospective revaluation is recovery from blocking. A recovery from blocking experiment includes three training phases followed by a test. In phase one of training, *A* is presented followed by *X*. In phase two of training, *AB* is presented followed by *X*. In phase three 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*-*X* relationship is blocked following phase two 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 three of the procedure) can become a conditioned exciter rather than a conditioned inhibitor of a presented outcome (as we demonstrated in our simulation of backward blocking). If Minerva-AL is to serve as a competent account of retrospective revaluation, it must handle recovery from blocking.

We applied Minerva-AL to the recovery from blocking procedure. The procedure had three successive training phases followed by a test. In phase one of training, *A* was presented followed by *X*. In phase two of training, the compound cue, *AB*, was presented followed by *X*. In phase three of training, *A* was presented alone. Following phase three, retrieval of *X* given *B* was tested. The design also included two control conditions. All three of the procedures are described schematically in Table 3. If Minerva-AL accommodates recovery from blocking, we should observe two results. First, retrieval of *X* given *B* should be greater than zero (i.e., $X|B > 0$) in the recovery from blocking condition. Second, retrieval of *X* given *B* should be reliably more positive in the recovery from blocking condition than in either of the control conditions.

We conducted 25 independent replications for each of the three conditions in Table 4 at each of three levels of the encoding parameter *L*. As shown in Table 4, 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 covaried with *L*.

Our explanation of recovery from blocking follows from the dynamics of storage and retrieval in Minerva-AL. Phase one established *A* as a retrieval cue of *X*. This learning caused blocking of the *B*-*X* relationship in phase two (see our prior simulation of blocking in Table 1). In phase three, *A* retrieved both *B* and *X* into the echo. Because neither *B* nor *X* were presented, memory recorded *A* paired with inverse representations of both *B* and *X*. At test, *B* retrieved the phase three traces (i.e., the $+A$, $-B$, $-X$ traces). Because traces are multiplied by their activations at retrieval, the $+A$, $-B$, $-X$ traces from phase three were reinverted at retrieval. Summing the reinverted traces produced a positive representation of *X* in the echo. Thus, according to Minerva-AL, recovery from blocking reflects new learning that *B* predicts *X* over phase three of the training procedure.

Backward conditioned inhibition. We have showed that Minerva-AL handles backward blocking and recovery from block-

Table 3
Simulation of Recovery From Blocking: Retrieval of X Given B as a Function of L (Standard Errors in Parentheses)

	Training				Learning rate		
Condition	Phase 1	Phase 2	Phase 3	Test	0.33	0.67	1.00
Recovery	50 $A \rightarrow X$	50 $AB \rightarrow X$	*50 A*	$X\|B$	.38 (.03)	.57 (.04)	.95 (.02)
Control (1)	50 $A \rightarrow X$	50 $AB \rightarrow X$	*50 C*	$X\|B$	.26 (.02)	.19 (.02)	.15 (.02)
Control (2)	50 $A \rightarrow X$	50 $AB \rightarrow X$		$X\|B$	.27 (.02)	.24 (.02)	.16 (.01)

Note. Means and standard errors are computed from 25 independent replications of the procedure. Numbers next to cues denote number of trials.

ing. We now turn to a third illustration of retrospective revaluation: backward conditioned inhibition.

Backward conditioned inhibition ("backward inhibition") is another example of retrospective revaluation. In a backward inhibition procedure, the learner is presented with pairings of a compound cue, *AB*, without an outcome. In a subsequent training phase, one element of the compound (i.e., *A*) is paired with an outcome, *X*. Following phase two of training, the learner behaves as though *B* signals the impending absence of *X*. That is, *B* behaves as a conditioned inhibitor of *X*.

The 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 an unpresented but associatively activated cue can develop or modulate its association to a presented outcome.

We applied Minerva-AL to the backward inhibition procedure. The procedure included two successive training phases followed by a test. In phase one of training, compound cue *AB* was presented followed by nothing. In phase two of training, cue *A* was presented followed by *X*. Following phase two, retrieval of *X* given *B* was tested. The design also included two control conditions (see Table 4). If Minerva-AL accommodates backward inhibition, we will observe two results at test. First, in the backward inhibition condition, retrieval of *X* given *B* should be less than zero (i.e., $X|B < 0$). Second, retrieval of *X* given *B* should be reliably more negative in the backward inhibition condition than in either of the control procedures (see Table 4).

We conducted 25 independent replications for each of the three conditions in Table 4 at each of three levels of the encoding parameter *L*. As shown in Table 4, 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* in the backward inhibition condition was less than retrieval of *X* given *B* in both of the control conditions. Finally, the size of the backward blocking effect increased as a function of *L*. In other simulations, we varied the value of *L* more broadly; the size of the effect diminished as a function of *L* however the model predicted backward conditioned inhibition in all cases.

Recently, Urcelay, Perelmuter, and Miller (2008) evaluated backward conditioned inhibition using a summation test. Their procedure included two successive training phases followed by a test. In phase one of training, *AB* was presented without an outcome. Phase two of training involved two intermixed kinds of trials. On half the trials, *A* was presented followed by *X*; on the remaining trials, *C* was presented followed by *X*. At test, retrieval of *X* given *C*, retrieval of *X* given *BC*, and retrieval of *X* given *CD* were tested. They reasoned that if *B* had become a conditioned inhibitor then $X|BC$ ought to be less than both $X|CD$ and $X|C$. The predictions were confirmed in the experiment. We applied Minerva-AL to Urcelay et al.'s procedure. Minerva-AL made the appropriate prediction: $X|BC < X|CD < X|C$. The result held across values of *L*.

Minerva-AL's explanation of backward conditioned inhibition is consistent with its explanation of backward blocking. In phase one of training, *A* was established as a retrieval cue for *B* (a within-compound association). Consequently, in phase two 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., $-B$) and a positive representation of *X*. At test, presenting *B* to memory caused those traces to invert (i.e., the $-B$, $+X$ traces were inverted as $+B$, $-X$ traces at retrieval). Summing the activated traces produced an inverse representation of *X* in the echo. Thus, like with backward blocking and recovery from blocking, Minerva-AL asserts backward conditioned inhibition is neither an encoding nor retrieval effect but rather falls out of an interaction between encoding and retrieval.

General Discussion

We adapted an instance-based model of human memory to simulate retrospective revaluation. In our account, memory records the events from individual trials. When a cue is presented to memory, it contacts all traces in parallel and causes each to become active. Each trace's activation is a positively accelerated function of its similarity to the probe. The information retrieved from memory, the echo, is a weighted sum of the activated traces. The model's anticipation that a target outcome is presented or withheld following presentation of a cue is indexed by comparing information retrieved in the echo against a target outcome. The ebb and flow of a cue's ability to retrieve an outcome from memory is the process of associative learning.

Minerva-AL accommodates acquisition, extinction, backward conditioning, blocking, conditioned inhibition, backward blocking, recovery from blocking, and backward conditioned inhibition—all by analogy to the process of cued-recall in human memory. Based on these successes, we argue that an instance-based theory of human memory that uses expectancy-encoding offers a coherent explanation of retrospective revaluation. In hindsight, Minerva-AL's facility with retrospective revaluation is unsurprising: retrospective revaluation involves a process of memory and Minerva-AL assumes learning is a memorial process (see Bouton & Moody, 2004, for a review of memory in learning).

There are at least five other computational accounts of retrospective revaluation. One account is based on Van Hamme and

Table 4
Simulation of Backward Conditioned Inhibition: Retrieval of X Given B as a Function of L (Standard Errors in Parentheses)

	Training			Learning rate		
Condition	Phase 1	Phase 2	Test	0.33	0.67	1.00
Backward inhibition	50 $AB \rightarrow$	50 $A \rightarrow X$	$X\|B$	−.48 (.02)	−.73 (.03)	−.86 (.01)
Control (1)	50 $AB \rightarrow$		$X\|B$	.00 (.01)	.01 (.01)	.00 (.01)
Control (2)	50 $AB \rightarrow$	50 $C \rightarrow X$	$X\|B$	.00 (.01)	−.01 (.01)	−.01 (.01)

Note. Means and standard errors are computed across 25 independent replications of the procedure. Numbers next to pairings denote number of trials.

Wasserman's (1994) modified Rescorla-Wagner account, one is based on Dickinson and Burke's (1996; see also Aitken & Dickinson, 2005) modified SOP model, one is based on Miller's comparator hypothesis (see Miller & Matzel, 1988; Miller & Schachtman, 1985; Stout & Miller, 2007), one is based on Ghirlanda's (2005) elemental model, and one is based on Daw and Courville's (2007) Bayesian model of learning.

According to Van Hamme and Wasserman (1994), retrospective revaluation follows from the formation of within-compound associations. To illustrate, consider backward inhibition. In phase one of training, *A* and *B* are presented together as cues. This establishes a within-compound association between *A* and *B*. Because of the within-compound association, *A* retrieves *B* in phase two of learning. To force conditioned inhibition, Van Hamme and Wasserman assign *B* a negative learning rate (i.e., reflecting the fact that *B* was expected but was not presented). Although it is not clear what a negative learning rate might mean, the explanation accommodates the result.

According to Dickinson and Burke's (1996) modified SOP model, stimuli are composed of elements that are in one of three states: a high activity state, a low activity state, or an inactive state. A stimulus with elements in the high activity state acquires positive associative strength to another stimulus with elements in the high activity state but acquires negative associative strength to a stimulus with elements in the low activity state. To explain backward conditioned inhibition, Dickinson and Burke propose that in phase two of training *A* retrieves both itself and *X* into the high activity state and retrieves *B* into the low activity state. Because, *B* is in the low activity state and *X* is in the high activity state, the two develop negative associative strength.

Miller and Schachtman's (1985) comparator hypothesis provides another account of retrospective revaluation. To illustrate, consider the problem of backward inhibition. According to the comparator model account of backward inhibition, when *B* is presented at test, it retrieves both a direct representation of *X* and an indirect representation of *X*: the indirect representation is retrieved using *A* as an intermediary (i.e., *B* retrieves *A* which retrieves *X*). Indirect representations are antagonistic to direct representations in the comparator framework and so the strength and direction of responding reflects the difference in associative strength between directly and indirectly retrieved representations. When the indirectly activated representation of *X* exceeds the strength of the directly activated representation of *X*, behaviour indicative of conditioned inhibition occurs. Because *B* is never paired with *X* in the backward inhibition procedure, the direct representation of *X* retrieved by *B* at test is comparatively weaker than the indirect representation of *X* given *B* that is retrieved indirectly through *A* (which was paired with *X* in training).

Ghirlanda (2005) developed an elemental account of retrospective revaluation. In his model, a stimulus is represented as a pattern of activity over a set of units, where *A* and *B* activate some units in common. Because *A* and *B* share units, presenting *A* is related to presenting *B*, and vice versa. Because *A* and *B* share units, presenting *A* leads to some learning about *B*, and vice versa.

Finally, Daw and Courville's (2007) Bayesian account of learning, in which they frame the process of learning to processes in a particle filter, provides yet another explanation of retrospective revaluation. To explain backward blocking, they argue that after learning that *AB* predicts *X*, learning that *A* alone predicts *X* causes an anticorrelated joint distribution over the weights connecting *A* to *X* and the weights connecting *B* to *X*. Because of the negative-correlation between *A*'s association to *X* and *B*'s association to *X*, learning that *A* predicts *X* forces the model to learn indirectly that *B* predicts the absence of *X*.

Ideally, one could evaluate the quality of Minerva-AL's explanation for retrospective revaluation relative to the explanations from other models. Unfortunately, no published data compel selecting one model over the others. There is, however, a way to critically evaluate our instance-based account.

Minerva-AL's unique explanation of retrospective revaluation leads to novel predictions for learning about associatively activated cues. To illustrate, consider an extension of the retrospective revaluation task that includes three successive learning phases. In phase one, *AB* is presented followed by *X*. In phase two, *BC* is presented followed by *Y*. In phase three, *CD* is presented followed by *Z*. In phase one, *A* becomes a retrieval cue for both *B* and *X*. In phase two, the within-compound association between *A* and *B* causes *BC* to retrieve a representation of *A*. Because *A* is retrieved but not presented, an inverse representation of *A* is stored in combination with positive representations of *B*, *C*, and *Y*. In phase three, *CD* retrieves the traces from phase two (including the inverse representations of *A*). Because *A* is retrieved into the echo, but *A* is not presented, the inverse representation of *A* is reinverted and memory for the trial records a positive representation of *A* paired with a positive representation of *Z*. Thus, following phase three and assuming no forgetting over trials, Minerva-AL predicts that *A* will be a conditioned exciter of *X*, a conditioned inhibitor of *Y*, and a conditioned exciter of *Z*. We are currently testing the prediction using a contingency judgement task.

Although we did not design Minerva-AL to do so, the model speaks to four debates in the study of associative learning. First, it has proven difficult for theories of associative learning to accommodate the recognition of unpresented cues. As we have already described, Minerva-AL recognised unpresented cues by encoding violations of its expectations. Second, early theories of associative learning described the growth of association between cues and outcomes. However, data indicate that associations grow between concurrently presented cues as well (i.e., within-compound associations). Minerva-AL learns within-compound associations. Third, there is a contentious debate on whether associative learning should be modelled as a process of encoding or a process of retrieval (see Miller & Escobar, 2001, for a discussion of the learning-performance distinction). Minerva-AL argues that learning reflects an interaction between encoding and retrieval. Fourth, a growing body of evidence for instance-like effects in learning has challenged traditional learning theories that do not represent instances. Minerva-AL acknowledges the growing body of evidence (e.g., Fagot & Cook, 2006; Griffiths, Dickinson, & Clayton, 1999; Hare & Atkins, 2001; Karte-Teke, De Souza Silva, Huston, & Dere, 2006).

Despite Minerva-AL's successes, the theory has limitations. One limitation follows from the scope of our demonstrations. Based on the work presented here, we claim that Minerva-AL stands as a competent account of retrospective revaluation. However, we do not claim Minerva-AL stands as a general theory of associative learning. To position Minerva-AL as a competitive and general theory of associative learning, we would be need to show it handles a broader array of learning problems, such as renewal,

stimulus generalization, stimulus discrimination, external inhibition, superconditioning, latent inhibition, overexpectation, recovery from overexpectation, overshadowing, recovery from overshadowing, and discrimination of cues presented singly and in compound. We are working on generalizing the model to these other protocols (Jamieson, Crump, & Hannah, 2010).

Another limitation of our account is its rudimentary representation of the timed presentation of cues and outcomes within a learning trial. In the simulations we have conducted, we present the cue (or cues) as a probe to memory, retrieve the echo, and then encode the events of the trial against the full event vector. Thus, whereas Minerva-AL can simulate a situation in which the cue (or cues) precedes the outcome (e.g., the distinction between forward and backward conditioning), Minerva-AL does not appreciate the difference between more subtle temporal manipulations, such as the distinction between delay and trace conditioning. Modelling the details of timed presentation on learning has been dealt with elsewhere in both the conditioning (e.g., Dickinson & Burke, 1996; Gallistel & Gibbon, 2000; Sutton & Barto, 1981; Wagner, 1981) and human memory literatures (e.g., Brown, Preece, & Hulme, 2000; Howard & Kahana, 2002). Developing a mechanism to simulate details of stimulus timing within the trial presents a challenge for future work.

Minerva 2 is one of several instance-based theories of human memory. So, then, why did we choose to build our account based on Minerva 2? We used Minerva 2 rather than a different theory for two reasons. First, the theory is simple and provides a clear method to discuss how storage and retrieval of instances can predict associative learning. Second, and most importantly, the Minerva 2 model is important in the study of human memory because it explains memory of the general based on memory for the particular. The problem of associative learning poses the same difficulty for an exemplar account: it is easy to imagine that memory for instances is at play in learning, but it is not so clear how memory for the generalities between cues and outcomes might emerge from the store of instances. By using Minerva 2, we have illustrated how contingency learning emerges out of the storage and retrieval of instances, without requiring a second learning system to compile contingency information. We do not wish, however, to suggest that an instance-based theory of learning must take the exact framework of Minerva 2. There are a number of instance-based models available (e.g., Kruschke, 1992; Vokey & Higham, 2004) that could achieve similar outcomes from different assumptions.

Minerva-AL is an adaptation of the Minerva 2 model. Because of that relationship, one might jump to the conclusion that Minerva-AL extends the reach of the Minerva 2 model to the domain of associative learning. That conclusion would be in error. Despite sharing common principles, the Minerva 2 and Minerva-AL models differ in measurable ways. We emphasise that Minerva 2 does not predict associative learning and that Minerva-AL can, in some cases, contradict predictions made from its parent theory. To integrate the two theories, one might speculate that expectancy-encoding operates only in cases where learning is unintentional or that expectancy-encoding is suppressed in situations where learning is deliberate. Until we can specify when expectancy-encoding should be at work and until we can specify a computational mechanism to control when expectancy-encoding will influence learning, we remain silent on the point. The integration of the two theories stands as a challenge for future work.

Résumé

Nous adaptons un modèle d'instance de la mémoire humaine, Minerva 2, afin de simuler la réévaluation rétrospective. Dans ce modèle, la mémoire préserve les événements des essais individuels sous forme de traces séparées. Une cible présentée à la mémoire contacte toutes les traces en parallèle et active chacune d'entre elles. L'information récupérée en mémoire est la somme des traces activées. L'apprentissage est modélisé comme un processus de rappel indicé; l'encodage est modélisé comme un processus d'encodage différentiel de caractéristiques inattendues de la cible (c.-à-d., encodage des attentes). Le modèle s'applique à trois exemples de réévaluation rétrospective : le blocage inversé, la récupération du blocage et l'inhibition conditionnée inversée. Ces travaux intègrent une compréhension de la mémoire humaine et de l'apprentissage associatif complexe.

Mots-clés : Théorie de l'instance, apprentissage associatif, réévaluation rétrospective, Minerva 2

References

Aitken, M. R. F., & Dickinson, A. (2005). Simulations of a modified SOP model applied to retrospective revaluation of human causal learning. *Learning and Behaviour, 33,* 147–159.

Arndt, J., & Hirshman, E. (1998). True and false recognition in MINERVA 2: Explanations from a global matching perspective. *Journal of Memory and Language, 39,* 371–391.

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.

Blaisdell, A. P., Gunther, L. M., & Miller, R. R. (1999). Recovery from blocking achieved by extinguishing the blocking CS. *Animal Learning & Behavior, 27,* 63–76.

Blaser, R. E., Couvillon, P. A., & Bitterman, M. E. (2004). Backward blocking in honeybees. *The Quarterly Journal of Experimental Psychology, 57*B, 349–360.

Bouton, M. E., & Moody, E. W. (2004). Memory processes in classical conditioning. *Neuroscience and Biobehavioural 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, NJ: Erlbaum Associations, 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, England: Cambridge University Press.

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review, 107,* 127–181.

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.

Daw, N. D., & Courville, A. C. (2007). The pigeon as particle filter. In J. Platt, D. Koller, Y. Singer, S. Roweis (Eds.), *Advances in neural information processing systems 20* (pp. 369-376). Cambridge, MA: MIT Press.

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

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.

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, USA, 103*(46), 17564–17567.

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

Ghirlanda, S. (2005). Retrospective revaluation as simple associative learning. *Journal of Experimental Psychology: Animal Behavior Processes, 31,* 107–111.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105,* 251–279.

Griffiths, D. P., Dickinson, A., & Clayton, N. S. (1999). Declarative and episodic memory: What can animals remember about their past? *Trends in Cognitive Science, 3,* 74–80.

Hare, J. F., & Atkins, B. A. (2001). The squirrel that cried wolf: Reliability detection by juvenile Richardson's ground squirrels (*Spermophilus richardsonii*). *Behavioral Ecology & Sociobiology, 51,* 108–112.

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behaviour Research Methods, Instruments, & Computers, 16,* 96–101.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411–428.

Hintzman, D. L. (1987). Recognition and recall in MINERVA-2: Analysis of the "recognition failure paradigm. In P. E. Morris (Ed.), *Modelling cognition* (pp. 215–229). London, England: Wiley.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95,* 528–551.

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology, 46,* 269–299.

Jamieson, R. K., Crump, J. C. M., & Hannah, S. D. (2010). An instance-based account of associative learning. Manuscript submitted for publication.

Jamieson, R. K., Holmes, S., & Mewhort, D. J. K. (in press). 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.*

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 on individual items. *Quarterly Journal of Experimental Psychology, 63,* 1014–1039.

Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. In B. A. Campbell & R. M. Church (Eds.), *Punishment* (pp. 279–296). New York, NY: Appleton-Century-Crofts.

Kart-Teke, E., De Souza Silva, M. A., Huston, J. P., & Dere, E. (2006). Wistar rats show episodic-like memory for unique experiences. *Neurobiology of Learning and Memory, 85,* 173–182.

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.

Kwantes, P. J. (2005). Using context to build semantics. *Psychonomic Bulletin & 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.

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. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review, 82,* 276–298.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85,* 207–238.

Melchers, K. G., Lachnit, H., & Shanks, D. R. (2004). Within-compound associations in retrospective revaluation and indirect learning: A challenge for comparator theory. *Quarterly Journal of Experimental Psychology, 57,* 25–53.

Miller, R. R., & Escobar, M. (2001). Contrasting acquisition-focused and performance-focused models of acquired behavior. *Current Directions in Psychological Science, 10,* 141–145.

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., & 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.

Miller, R. R., & Schachtman, T. R. (1985). Conditioning context as an associative baseline: Implications for response generation and the nature of conditioned inhibition. In R. R. Miller & N. E. Spear (Eds.), *Information processing in animals: Conditioned inhibition* (pp. 51–88). Hillsdale, NJ: Erlbaum.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115,* 39–57.

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not unconditioned stimuli. *Psychological Review, 82,* 532–552.

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

Shanks, D. R. (1985). Forward and backward blocking in human contingency judgment. *Quarterly Journal of Experimental Psychology, 37,* 1–21.

Stout, S. C., & Miller, R. R. (2007). Sometimes-competing retrieval (SOCR): A formalization of the comparator hypothesis. *Psychological Review, 114,* 759–783.

Sutton, R. S., & Barto, A. G. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review, 88,* 135–170.

Urcelay, G. P., Perelmuter, O., & Miller, R. R. (2008). Pavlovian backward conditioned inhibition in humans: Summation and retardation tests. *Behavioural 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 & Motivation, 25,* 127–151.

Vokey, J. R., & Higham, P. A. (2004). Opposition logic and neural network models in artificial grammar learning. *Consciousness and Cognition, 13,* 565–578.

Von Restorff, H. (1933). Uber die Wirkung von Bereichsbildungen im Sperunfeld [On the effect of spheres formations in the trace field]. *Psychologische Forschung, 18,* 448–456.

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

Whittlesea, B. W. A., & Williams, L. D. (2000). The source of feelings of familiarity: The discrepancy-attribution hypothesis. *Journal of Experimental Psychology: Learning, Memory, & 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.

Received December 4, 2009
Accepted July 8, 2010 ■