

Differentiating Models of Associative Learning: Reorientation, Superconditioning, and the Role of Inhibition

Brian Dupuis and Michael R. W. Dawson
University of Alberta

A recent associative model (Miller, N.Y., & Shettleworth, S.J., 2007. Learning about environmental geometry: An associative model. *Journal of Experimental Psychology: Animal Behavior Processes B*, 33, 191–212) is an influential mathematical account of how agents behave when reorienting to previously learned locations in spatial arenas. However, it is mathematically and empirically flawed. The current article explores these flaws, including its inability to properly predict geometric superconditioning. We trace the flaws to the model's mathematical structure and how it handles inhibition. We then propose an operant artificial neural network model that solves these problems with inhibition and can correctly model both reorientation and superconditioning.

Keywords: Rescorla–Wagner model, artificial neural networks, operant choice, reorientation, superconditioning

Perhaps the most influential formal account of associative learning is the classic Rescorla–Wagner model (Rescorla & Wagner, 1972):

$$\Delta V = \alpha \cdot \beta \cdot (\lambda - \sum V) \quad (1)$$

In the Rescorla–Wagner model, the change in associative strength ΔV between conditioned stimulus (conditional stimulus, CS) and unconditioned stimulus (US) is defined by the difference between the magnitude of the US, represented by λ , and magnitude of the current associative strength, represented by V . This difference is scaled by the CS' inherent salience α and by the learning rate related to the US, β . This model is inherently one of *classical* conditioning. At every iteration, the equation updates associative strengths for all presented cues simultaneously; the agent's choice to respond (or not) is not part of this formulation of learning (Dawson, 2008).

Recently, there have been attempts to extend the Rescorla–Wagner model to handle other types of conditioning. For example, Miller and Shettleworth (2007) convincingly argue that a case in which the agent only chooses one option at a time from a selection of possibilities, and only receives reinforcement or feedback about the chosen option, would be more properly considered an *operant* task, with reinforcement contingencies based upon the agent's

particular pattern of choices. Geometric reorientation (Cheng, 1986) is presented as an example of this type of task, in which disoriented agents choose to search particular locations of a controlled arena for reinforcement. (This task is described in greater detail below.) Under an associative viewpoint, locations of interest within these arenas are viewed as collections of cues, and each cue competes for associative strength. In attempting to model geometric reorientation using a more “operant” view of association, Miller and Shettleworth modify Equation 1 to include a measure of the agent's probability of choosing a given location:

$$\Delta V_E = \alpha_E \cdot \beta_L \cdot (\lambda_L - V_L) \cdot P_L \quad (2)$$

Here, the change in associative strength of each cue (or “element,” E) is updated using the sum of the associative strengths of all cues at a given location (i.e., $V_L = (\sum V_{E/L})$, and scaled by a term representing the probability of choosing that location. In its original version (Miller & Shettleworth, 2007), this model defined the probability of choosing a given location P_L as the relative associative strength of the location in question compared with the total associative strength at every possible location:

$$P_L = \frac{V_L}{\sum V_L} \quad (3)$$

However, Dawson, Kelly, Spetch, and Dupuis (2008) identified a serious problem when probability is defined using Equation 3. They demonstrated that, using an example reorientation problem taken from Miller and Shettleworth (2007), a model that used Equations 2 and 3 can produce values of P_L that fall outside the range of 0 to 1, and thus, cannot be considered “probabilities.” In response to the flaw identified by Dawson et al. (2008); Miller and Shettleworth (2008) revised their model of reorientation. Miller and Shettleworth replaced Equation 3 with a new term for the relative net *attractiveness* of a location. In the modified model, Miller and Shettleworth defined “net attractiveness of a location” r_L as the sum of the associative strengths of the cues at that location if that sum is positive, or as 0 if that sum is not positive

This article was published Online First April 29, 2013.

Brian Dupuis and Michael R. W. Dawson, Department of Psychology, University of Alberta, Edmonton, Alberta, Canada.

This work is based on part of a master's thesis defended in September, 2012, for the Department of Psychology, University of Alberta. We thank Marcia Spetch and Patricia Boechler for their aid, as well as Nolan Thomas Denman, Gerry Leenders, Michel Fiallo-Perez, and Vadim Kasim for helpful discussion during preparation of the article.

Correspondence concerning this article should be addressed to Brian Dupuis, Department of Psychology, University of Alberta, Edmonton, AB, T6G 2E9. E-mail: bdupuis@ualberta.ca

(that is, $r_L = V_L * H_0(V_L)$, in which H_0 is the Heaviside step function with threshold 0). P_L became the relative net attractiveness of each location compared with the total relative net attractiveness at every location:

$$P_L = \frac{r_L}{\sum r_L} \quad (4)$$

They reasoned that by replacing Equation 3 with Equation 4, P_L will always fall within the acceptable range for probability, and they presumed that the theoretical issues with their model had been resolved.

Below, we show that even after this modification, the model is still inadequate for modeling learning in the reorientation task or related associative tasks. The purpose of the current article is to demonstrate particular inadequacies empirically, and to explain their source in the model's equations. This analysis of their revised model will also reveal why the errant behavior identified by Dawson, et al. (2008) emerged in the original version of Miller and Shettleworth's model. We analyze and explain the model's behavior—and errors—in two separate spatial-learning tasks: spatial reorientation and geometric superconditioning. Finally, we propose an alternative model to rectify the situation.

Although Miller and Shettleworth (2007) also present a “multiple-choice” version of their model in which agents are allowed to search multiple locations before receiving feedback, the mathematics of this are an extension of the single-choice model, and therefore any mathematical flaw in the single-choice model is inherited in the multiple-choice model as well. For this reason, we focus on the single-choice version below. By convention, we will refer to models with the form of Equation 2 as the “M–S model” regardless of the equation used to compute P_L . The model's original presentation (Miller & Shettleworth, 2007), consisting of Equations 2 and 3, is referred to as “M–S 2007,” and the revised form (Miller & Shettleworth, 2008) using Equations 2 and 4 is denoted as “M–S 2008.” The correction employed by replacing Equation 3 with Equation 4 is referred to as the “positiveness correction,” because unlike V_L , r_L is defined in such a way that it is always positive, and never negative (as negative sums are artificially set to 0).

Reorientation

The Task

The Miller–Shettleworth associative model was designed around a particular case study in association: reorientation, defined here as an agent's ability to locate a previously learned position when disoriented. At its simplest, reorientation requires an agent to navigate to a previously learned location based upon available environmental cues, such as information about the shape of the environment (“geometric” cues) and about landmarks present in the environment (“feature” cues). Such behavior is typically studied experimentally with what has become known as the “reorientation task” (Cheng, 1986), which has shown that agents, even those from dramatically different species that range from ants to humans, exhibit certain empirical regularities (reviewed in Cheng & Newcombe, 2005).

In a typical reorientation task (Cheng, 1986), an agent freely explores a rectangular arena, with each corner distinguished from

the others through some combination of geometric and featural cues. If the agent approaches the corner the experimenter has deemed “correct”—for instance, the corner containing a unique feature like a colored panel, with a long wall on its left—then the agent is reinforced, otherwise no reward is offered. After repeated trials to learn that this location is correct, the agent is disoriented and placed in a new arena in which the feature cues are placed in conflict with the geometric cues—continuing the earlier example, the panel is now in a corner with a long wall on its right. It might be plausible to predict that, in this new arena, agents will move toward the feature, which was the only unique predictor of reward during training. Curiously, although agents will approach the corner with the unique local feature, agents will also frequently choose locations matching the original geometry—the original corner and its geometric equivalent—even though this geometry was not always reinforced during training, and neither corner currently possesses the reinforced local feature. The exact proportion of responses that follow the feature as opposed to the geometry varies somewhat depending on the size of the arena and the agent in question (e.g., see Cheng & Newcombe, 2005; Chiandetti & Vallortigara, 2008), but the general pattern remains consistent. A typical reorientation arena is depicted in Figure 1.

Figure 1 illustrates a training arena employed in a typical reorientation task (Wall, Botly, Black, & Shettleworth, 2004, Experiment 3). This experiment is of interest because it is the case study in which Miller and Shettleworth's (2007) associative model was developed and tested. Within this task, rats are required to locate food in one corner of a geometric arena—here, the Correct corner, with a particular set of geometric properties G and a unique feature F (along with a general context cue which Miller and Shettleworth labeled B , representing the bowls at each location that the rats searched for food). During this phase of exploration and learning, the rats are also exposed to a different set of geometric properties learned to be wrong, W , which never contain reinforcement. In later phases of this experiment, the configuration of cues changes (for instance, F might move from the Correct location to

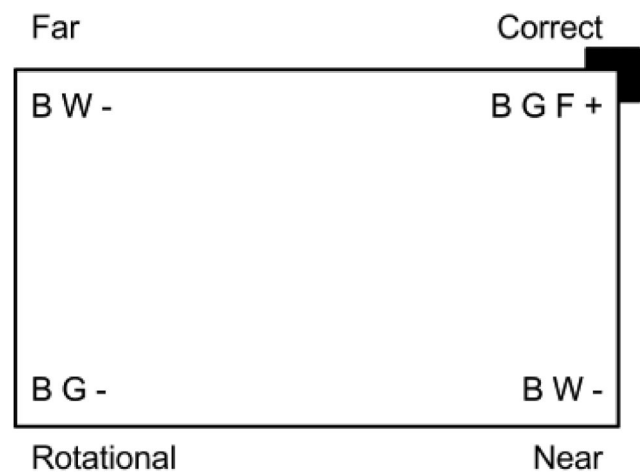


Figure 1. Schematic overview of the Wall et al. (2004) reorientation task training phase, traditionally used in discussions of the M–S model, with locations and cue types labeled. B = “Bowl” (a general context cue), G = “Correct Geometry”, F = “Feature”, W = “Wrong Geometry.” A+ indicates the location is reinforced, and a– indicates no reinforcement.

the Near location) and the rats' behavior is monitored, but for the purposes of the current article, we concentrate on this training phase.

Demonstration of Problems

The M–S 2008 model was created to solve problems that emerged when the M–S 2007 model simulated the training phase of this paradigm (Dawson, Kelly, Spetch, & Dupuis, 2008). In particular, the positiveness correction in M–S 2008 was intended to prevent the impossible probabilities M–S 2007 generated with high $\alpha\beta$ or over extended runs. Miller and Shettleworth (2008) found that when their revised model was used for the Wall, Botly, Black, and Shettleworth (2004) paradigm, it behaved correctly, even when high values (0.65) for $\alpha\beta$ were used. However, we show below that when some of the parameters of the revised model are slightly changed, errors are still produced. In short, the positiveness correction has not created a proper mathematical model of reorientation.

For example, when used to model the Wall et al. (2004) reorientation task at slightly higher learning-rate values than $\alpha\beta = 0.65$, the M–S 2008 model predicts dramatic *fluctuations* in associative strength and choice probabilities when both should instead plateau, suggesting a lack of robustness in the underlying mathe-

matics. (An analysis of the mathematical properties and problems of the M–S model is provided in a later Appendix.) The fluctuations in associative strength escalate until they eventually culminate in a global divide-by-zero error. This is repeatable and can be predicted reliably; for instance, for this task at $\alpha\beta = 0.7$, the fluctuations begin between iterations 10 and 20 (depending on the cue or location) and the collapse occurs at iteration 183 (see Figure 2). Similar fluctuations, in addition to the singularities reported by Dawson et al. (2008), can be demonstrated in the M–S 2007 formulation as well.

Why Does It Fail?

When modeling the Wall et al. (2004) reorientation task, the Miller and Shettleworth (2008) model produces two distinct errors, both illustrated in Figure 2. The first is a series of fluctuations of both associative strengths and choice probabilities at high learning rate $\alpha\beta$; similar behavior is observed in the Miller and Shettleworth (2007) model as well, suggesting that this problem's root cause is shared by both models. The second problem, in which the equations consistently produce a divide-by-zero error after several fluctuations, is unique to the M–S 2008 model, suggesting it is a consequence of the positiveness correction. We describe the source of these problems using terminology consistent with Miller and

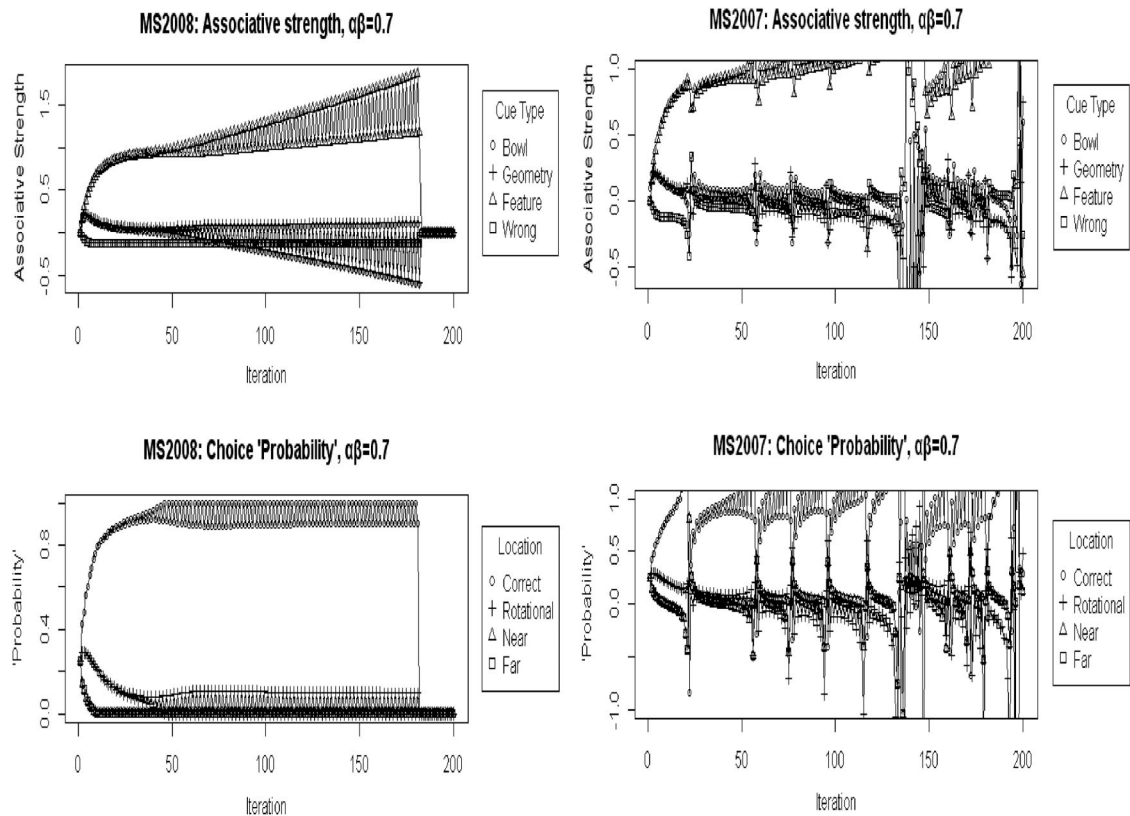


Figure 2. The two Miller–Shettleworth models (2007, 2008) simulating the Wall et al. (2004) reorientation task at a high $\alpha\beta$. Note the fluctuation in both models. For the M–S 2008 model (left), the 0s in the lower panel are actual 0 values, and the upper panel uses 0 as a replacement for divide-by-zero, occurring after the crash at iteration 183. The singularities present in the M–S 2007 model (right) were reported by Dawson et al. (2008), and the fluctuations are novel.

Shettleworth's (2007) discussion of reorientation, in which the task consists of four locations (Correct, Rotational, Near, and Far) that are defined in the model as combinations of four cues [a general context cue (B), Feature (F), correct Geometry (G), and Wrong geometry (W)]. G is present at the Correct and Rotational locations, and the Near and Far locations have the W cue. The B cue is present at all four locations, and the F cue is present only at the Correct location (see Figure 1). During simulation, all cues are initialized to 0 associative strength, except for B , which is initialized to 0.1 to reflect prior experience acquired during pretraining.

The fluctuations emerge from the structure of Equation 2, in which the change in associative strength for a cue is scaled by a function of associative strength (Equation 3 or Equation 4). Both of these scaling functions can approach zero when considering locations with strong inhibitory cues, causing the model to reduce the effective change in associative strength due to lack of reinforcement relative to the effective change in associative strength due to reinforcement. That is to say, the more inhibitory a location's cues become, the less learning takes place at that location relative to others. Because these locations contain cues which are at least partially reinforced on other locations (the B and G cues in this example), these cues acquire a greater positive change in associative weight from reinforced locations and a lesser negative change in associative weight from nonreinforced locations. At low learning rates $\alpha\beta$, this artificial inflation is small relative to the net weight, and easily handled by the error-correcting Equation 2. However, at sufficiently¹ high $\alpha\beta$, the artificial inflation is large enough to lead to an "overcorrection"—the magnitude of the change in weights being larger than it should be.

For example, consider the G cue—which is initialized at 0, is presented at two locations, and is only reinforced at one. Therefore, as reported by Miller and Shettleworth, we expect its associative strength to climb (as it is reinforced), peak (because this reinforcement is not universal), and stabilize at some small positive value. With $\alpha\beta = 0.7$, the peak occurs after the third iteration at $V_G = 0.22$; the corresponding $\Delta V_G = -0.01$ is negative, consistent with predictions—and this is expected to slowly decrease in magnitude with subsequent iterations. However, the subsequent iteration is not a decrease—rather, it increases, with $\Delta V_G = +0.003$. When evaluating the change in strength for the G cue, the scaling nature of Equations 3 and 4 produces a lower P_L for the nonreinforced Rotational corner than for the reinforced Correct corner—and as a result, the model assigns insufficient inhibitory strength to the G cue. Following this slight positive increase, the model "overcorrects" with a strong negative ΔV_G to reflect the Rotational corner—but because that corner contains the B cue as well as the G cue, the negative overcorrection applies to B as well. On subsequent iterations, these interlinked overcorrections produce the distinctive fluctuating behavior seen in Figure 2.

This fluctuation emerges from the structure of Equation 2, but the eventual "crash" seen in Figure 2 results from a divide-by-zero error in Equation 4. This occurs because Equation 4 still computes the "effective" V_L for every location in the model, and then multiplies that V_L by the Heaviside step function of V_L with threshold 0. This results in V_L if $V_L > 0$, and 0 otherwise—as locations' net associative strengths fall below zero, they are "dropped out" of the denominator of Equation 4. However, due to the fluctuations seen above, which grow in magnitude over time, a time will come in which the magnitude of the negative associative

strength of B and G together exceeds the magnitude of F , the only uniquely positive cue in the system. At this point, all four locations have negative net associative strength, and thus, Equation 4 attempts to divide by zero.

If Equation 4 is replaced with Equation 3 (resulting in the M-S 2007 model), a "crash" does not occur. Instead, the model produces singularities (identified by Dawson et al., 2008, illustrated in Figure 2). Without the positiveness correction, any location's net associative strength V_L is allowed to fall below zero—or, as noted above, for *all* locations' net associative strengths to fall below zero. At this point, the denominator in Equation 3 flips sign from positive to negative, and a singularity appears in the corresponding associative strengths and choice probabilities. As Equation 3 does not include an artificial substitution of 0, the divide-by-zero outcome does not happen, and if the model is allowed to continue to run, eventually the denominator of Equation 3 will become positive again, resulting in the next singularity in Figure 2, and so on.

Because these fluctuations and (in the M-S 2008 model) crashes arise after a different number of iterations depending on the chosen learning rate parameters $\alpha\beta$, and do not seem to appear during any reasonable span of time within specific ranges for $\alpha\beta$, it appears that a necessary step in applying the Miller–Shettleworth (2007, 2008) model is missing. Specifically, one must carry out a search of parameter space to find the boundaries in which the model will fail; such a search would need to be carried out for each permutation and combination of cues and locations present within the task. However, it is informative that points of failure exist at all: Such failures suggest an underlying problem with the mathematics of the model. At the end of this article we provide a Technical Appendix that reveals exactly what this underlying problem is.

The empirical failures demonstrated here appear when using the Miller–Shettleworth models to simulate reorientation—the task for which the models were developed. Other tasks can be viewed in an associative light as well, and therefore may see the Miller–Shettleworth model applied to them. Do these tasks also demonstrate empirical failures? We turn now to examine the Miller–Shettleworth models' performance on a different spatial task: that of geometric superconditioning.

Superconditioning

The Task

An interesting prediction of the Rescorla–Wagner model is that "superconditioning" can occur. Superconditioning exists when excitatory cues produce stronger responses after discrimination training if they are paired with an *inhibitory* cue, compared with a control condition in which the excitatory cues are paired with a neutral cue during training. The presence of an inhibitory cue during discrimination training increases the difference between λ and ΣV , allowing for a greater change in associative strength.

Horne and Pearce (2010, Experiment 2) investigated if one could observe superconditioning in the context of geometric and

¹ What qualifies as "sufficient" varies dramatically with the structure of the problem—cue distributions, number of locations, and initial associative strengths. For the Wall et al. (2004) task as described here, "sufficient" is near $\alpha\beta = 0.68$.

feature cues in a paradigm that can be viewed as an extension of the reorientation task. In this experiment (summarized in Figure 3), rats were trained to associate a particular set of geometric cues with reinforcement, but only when a particular feature was absent. That is, in Stage 2 of Figure 3, rats are reinforced when a particular set of cues are present at a location (A+), but are not reinforced when those cues are accompanied by an additional cue (AX-). Following this training, the experimental group of rats received reinforcement in the same location with both sets of cues present (AX+), and a control group received reinforcement in that location when the original cues are paired with a novel, neutral cue set (AY+).

Because of their attempt to model animal data using the M-S 2008 model, Horne and Pearce defined the task as consisting of a “correct geometry” cue G , an “incorrect (wrong) geometry” cue W , a context cue common to every location B , and two feature cues (one inhibitory feature present during discrimination training and with the experimental group, F , and one neutral feature only present for the control group, N)². A summary of these cues as presented in Horne and Pearce’s experiment is found in Figure 3.

Demonstration of Problems

Horne and Pearce’s (2010) animal results show that the experimental group chose the “correct” corner with greater frequency than did the control group—that is, the rats showed evidence of superconditioning. However, when simulating the same experiment with the M-S 2008 model, Horne and Pearce obtained the opposite result: The probability of choosing the correct corner in the experimental group was 0.92, and the probability of choosing the correct corner in the control group was 0.95. Although this difference is small, it is in the wrong direction relative to animal data, leading Horne and Pearce to conclude that the M-S model failed to display superconditioning. They attributed this to an artificial inflation in the value of the B cue, and attempted to solve the problem by manipulating the salience of this cue—in essence, performing a search of parameter space, as described above—but “in all of the simulations that [they] conducted, however, this manipulation does not permit superconditioning to be predicted” (p. 393). Such an outcome is a natural risk of systems that require parametric searching: It is possible that no viable parameter combinations exist that will solve the problem. Horne and Pearce appear to have reached this conclusion regarding the M-S models and superconditioning.

Why Does It Fail?

This artificial inflation of B is a symptom, but not the true cause, of the M-S model’s difficulty with superconditioning. Instead, the failure to capture Horne and Pearce’s (2010) superconditioning results stems from the structure of the M-S 2008 probability equation, which includes the “positiveness correction.” This term is employed to prevent the probability of choosing a location P_L from falling below zero even when that location’s net associative strength is below zero (Dawson et al., 2008). Horne and Pearce (2010) encounter such a location during Stage 2 (discrimination) of their simulation: The magnitude of the negative associative strength of the consistently inhibitory feature cue F far exceeds the magnitude of the positive associative strength of the other cues presented alongside it (B and G).

During Horne and Pearce’s (2010) Stage 2 (discrimination), the feature cue F is only present at a location which is consistently inhibited, causing F ’s associative strength to fall. Additionally, whenever F is inhibited, the other cues presented with it—the context cue B and the correct geometry cue G —are also inhibited, gaining a more negative associative strength. When the total associative strength at these locations— V_L , or the sum of the associative strengths of F , B , and G —falls below zero, the Miller-Shettleworth probability equation (Equation 4) applies the positiveness correction to set the corresponding P_L to zero, and the model should subsequently predict no further changes in associative strength in response to this location. A similar story can be told for the incorrect geometry cue W , which is never reinforced; once the sum of associative strengths at these locations ($B + W$) falls below zero, the positiveness correction prevents these cues from gaining further inhibitory strength.

However, the B and G cues are still present and reinforced at other locations on the acquisition trials during this stage, allowing those cues to continue to gain associative strength at a slow rate. These cues continue to gain positive associative strength from their reinforced presentations while the Miller-Shettleworth equation artificially prevents the cues from gaining negative associative strength during their inhibited presentations as soon as the net associative strength falls below zero. Concurrently, the associative strength of these cues (B and G) grows higher than it should. Similarly, during the acquisition trials, the B and G cues continue to gain positive associative strength from presentations at the reinforced locations, and the B , W , and G cues are all prevented from gaining adequate negative associative strength from nonreinforced locations due to the positiveness correction artificially setting such updates to 0.

During the Stage 3 (experimental or control group), this effect is inflated in the experimental group relative to the control group, as the experimental group pairs the already-inflated B and G cues with a now-consistently reinforced F cue (leading to higher effective P_L , and therefore a higher change in associative strength, on the correct locations). In contrast, although the control group pairs these cues with a never-reinforced novel N cue (with 0 initial associative strength), this does not inflate the change on the correct locations relative to the incorrect ones. Accordingly, the magnitude of the changes in associative strength will be greater for experimental groups than for control groups, even on locations that are identical between the two conditions. In effect, the model can acquire *more* inhibition from a nonreinforced corner during superconditioning than it could from the *same* nonreinforced corner during control.

This weakness in handling inhibition, combined with the overall scaling of all changes in associative strength, results in the model generating incorrect predictions as described by Horne and Pearce. For instance, the probability of the model selecting a correct corner during a test trial (without any feature cues) is given by the ratio

² Horne and Pearce (2010) employ a different notation, instead using G_C and G_I to refer to correct and incorrect geometric cues, and F to refer to any feature cue (despite two being used in the experiment). For consistent terminology across the two experiments, and to avoid confusion about whether a feature cue had prior associative strength (as it would have, during superconditioning) or if it is a novel, neutral cue (as it would be, during controls), we adopt the G , W , F , and N notation described here.

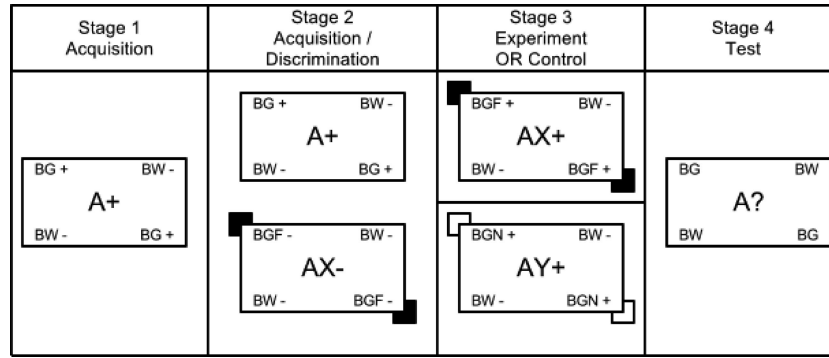


Figure 3. A schematic overview of the Horne and Pearce (2010) superconditioning task, adapted from that article, with stages and cue types labeled. *B* = “Base”, *G* = “Correct Geometry”, *W* = “Incorrect Geometry”, *F* = “Feature Cue”, *N* = “Neutral Feature Cue”. A + indicates the location is reinforced, a- indicates no reinforcement. During Stage 3, the upper panel depicts the experimental condition, and the lower depicts the control.

between the net associative strength at that corner and the total net associative strength, or $(B + G)/(B + G) + (B + W)$. All of these cues are artificially (and identically) inflated for both groups due to the errors during the second stage (discrimination), but during the third stage (experimental or control groups), the *W* cue has associative strength of -0.35 in the experimental group, a 65% increase relative to -0.21 in the control group. Accordingly, the probability of choosing a correct corner is higher in the control group (0.94) than the experimental group (0.91).

It is perhaps unsurprising that the source of these errors is not obvious at first glance. At a fundamental level, they emerge because the model represents locations as collections of cues, in which some of those cues are present at locations that are reinforced differently. This shared-cue perspective is not taken during Horne and Pearce’s (2010) discussion on superconditioning. Instead, they employ the common “A+/AX-” notation to discuss superconditioning—but in a cue competition perspective, “A” might represent some collection of cues that are not reinforced with equal likelihood. Indeed, here, “A” refers to *B* + *G*, which are present and reinforced in different proportions; similarly, the *W* cue (a vital part of the behavior described above) is simply not included in “A+/AX-” notation. As a consequence, Horne and Pearce focus on the feature cue *F* becoming a conditioned inhibitor dependent upon the context cue *B* becoming a conditioned excitor, but do not discuss that *B* is, in truth, both reinforced and not reinforced depending on the collection of cues present at a location.

Discussion

From this examination, it is now clear that the Miller and Shettleworth (2007, 2008) models are empirically flawed, and that these flaws stem from their mathematical formulation. Because similar errors occur with both the M–S 2007 and M–S 2008 model, it is unlikely that any flaws can be solely due to how the model defines probability (the source of the problems identified by Dawson et al., 2008). Instead, the errors discussed above emerge due to the structure of Equation 2 itself; merely altering the P_L equation but retaining Equation 2 will not solve this problem. (This assertion is discussed with greater rigor in the Technical Appendix.)

These results suggest that the Miller–Shettleworth (2007, 2008) model be abandoned for future research.

However, in spite of these empirical flaws, Miller and Shettleworth (2007) make an important observation concerning the application of associative models to geometry learning tasks: The agent’s pattern of behavior alters each cue’s apparent reinforcement contingencies on any given trial. In short, Miller and Shettleworth note that such learning is intrinsically *operant*. Miller and Shettleworth endeavored to model this by using the P_L equation to scale the Rescorla–Wagner model as formalized in Equation 2. Erroneously, this scaling resulted in the improper handling of net negative associative weights, producing the problems that have been described above. Importantly, the choice of P_L equation—Equation 3 or Equation 4—merely alters the form these issues take: It is within the structure of Equation 2 that the true problem lies.

The question now becomes if this is a problem with associationism in general—if an associative model can be created which makes use of Miller and Shettleworth’s (2007) operant observation, but is not subject to the empirical errors present in their model. In the next section, we describe an alternative model that has just these properties: it solves the empirical problems with Miller and Shettleworth’s model, and is still both associative and operant in nature.

A Solution to the Problems

Logistic Behavior

To begin, let us consider Miller and Shettleworth’s (2007) view of operant learning as follows: At some moment in time, an agent perceives a set of cues V_L related to a particular location L . The agent uses these cues to make a judgment about how attractive this location is. For instance, the agent might use these cues to predict the probability P_L of being reinforced if that location is actually visited. Indeed, we could also say that P_L is the likelihood that the agent will actually visit location L . If the location is visited, then the agent will be reinforced (or not), and can modify the associative strengths of the available cues accordingly. Such learning is

operant, because associative strengths will only be modified if the agent explores the location.

We saw earlier that one source of the problems with the M–S models is that either equation used to model P_L (i.e., either Equation 3 or Equation 4) has problems when faced with inhibition. Clearly, we need to select a different equation for P_L , one that is more robust to the negative associative strengths of inhibitory cues. A natural choice for this equation is the logistic function (given in Equation 5), which produces a response between 0 and 1 for all possible input values and monotonically increases as input increases.

$$P_L = \frac{1}{1 + e^{(-\sum(V_E)_L)}} \quad (5)$$

It is impossible for the logistic function to produce a value outside of the range between 0 and 1 (solving a problem that occurs with Equation 3), or to result in a divide-by-zero error (solving a problem that occurs with Equation 4). The logistic function is an ideal choice for computing probability like P_L (Dawson & Dupuis, 2012), and has a long history of being used to model phenomena in a wide variety of disciplines (Cramer, 2003).

The Perceptron

We also noted earlier that another source of the problems with the M–S models arose when their equations P_L were placed in the context of the remainder of Equation 2³. Importantly, the logistic equation permits us to take advantage of a different formalism that eliminates this difficulty. Equation 5—which converts weighted cues into a probability—defines a modern version of a very simple artificial neural network, called a perceptron (Rosenblatt, 1958, 1962). The simplest version of a modern perceptron (Dawson, 2004, 2008) consists of a set of input units, each of which can be used to represent whether a particular cue is present or absent. Each of these input units can send a signal to a single output unit through a weighted connection; the weight of the connection represents the associative strength of a particular cue. The output unit works by summing the weighted signals from the input units to produce a single number, called net input, which is identical to $(\sum V_E)_L$. The output unit then produces a response—its activation—by computing the logistic function of its net input exactly as defined by Equation 5.

Although Equation 5 defines the activation of this perceptron, it does not define the learning rule with which the connection weights are adjusted. Such a rule can be found in (Dawson, 2008), expressed here with the language of Miller and Shettleworth:

$$\Delta V_E = \alpha_E \beta_L \cdot (\lambda_L [US_L] - P_L) \cdot [CS_E] \quad (6)$$

Here α , β , and λ are used as they are in Equation 1 and 2, while “ US_L ” reflects presence (1) or absence (0) of reinforcement (the unconditioned stimulus) at a location, “ CS_E ” reflects the value of the input unit corresponding to element E (also 1 or 0, for presence or absence of the conditioned stimulus), and P_L reflects the perceptron’s logistic response to a given pattern of cues, given by Equation 5 (a function of $(\sum V_E)_L$).

In employing Equations 5 and 6, we are in essence proposing that the perceptron can provide a central component of an associationist model of spatial learning. There are several reasons that this proposal is attractive.

First, the mathematics of this type of model are well-established—there is a long history of mathematical results concerning perceptron learning, beginning with the work of Rosenblatt (1958, 1962).

Second, the associative models of Miller and Shettleworth (2007, 2008) are extensions of the well-established Rescorla–Wagner model of associative learning (Rescorla & Wagner, 1972). Importantly, the kind of learning that is carried out by a perceptron can be formally translated into Rescorla–Wagner learning (Dawson, 2008; Gluck & Bower, 1988; Sutton & Barto, 1981).

Third, the motivation behind Equation 5 was to generate a value that could be interpreted as a probability. It has been shown empirically that perceptron responses can be interpreted as probabilities, because these networks can learn to generate responses that match the probabilities of events occurring in the world (Dawson, Dupuis, Spetch, & Kelly, 2009). Furthermore, formal analyses of perceptrons prove that the activity of an output unit can literally be called a conditional probability (e.g., Dawson & Dupuis, 2012).

Fourth, one of the reasons that the M–S models are of interest is because they have been argued to be able to model reorientation task regularities. Crucially, perceptrons have also been shown to be capable of modeling a variety of reorientation task phenomena (Dawson, Kelly, Spetch, & Dupuis, 2010). These include successful modeling of basic reorientation (feature and geometric cue processing, with appropriate errors) in rectangular and quadrilateral arenas, reorientation with multiple unique landmark cues in assorted transformations, and evidence for changing emphasis on geometric cues as arena size is varied. The perceptron can handle other associative regularities not specific to reorientation as well, but these fall beyond the scope of the current article; the interested reader is referred to Dawson (2008) for an overview.

An Operant Algorithm

Although the perceptron has been successfully used to model many empirical regularities found in the reorientation task (Dawson et al., 2010), this was done using training procedures that are *not operant* in nature. We now describe an algorithm which trains a perceptron in an operant fashion, transporting Miller and Shettleworth’s (2007) core idea about reorientation into the domain of artificial neural networks.

The typical, and nonoperant, manner for training a perceptron (Dawson, 2004, 2008) proceeds as follows: First, a pattern (i.e., a set of cues, such as those corresponding to one “location”) is presented to the perceptron’s input units. Second, the perceptron converts input unit signals into an output response (i.e., Equation 5). Third, the perceptron receives feedback about its response (e.g., it receives reinforcement, or not). Fourth, a learning rule (Equation 6) is used to modify connection weights (associative strengths) in accordance with the feedback.

A simple change to the above learning algorithm makes it truly operant (Dawson et al., 2009). The second step in the above procedure is to compute P_L using Equation 5. Once this is computed, we can add a new step in which P_L is used to make a

³ Simply substituting the logistic equation for P_L in Equation 2 will not solve the problem; for a more formal discussion of these problems and their consequences, we refer the interested reader to the Technical Appendix.

choice—in essence, a choice about whether or not to visit location L —in which the likelihood of visiting the location is P_L . If the choice is made to visit the location, then learning proceeds according to the third and fourth steps in the above algorithm. However, if the choice is made to *not* visit the location, then no learning occurs—connection weights are not updated, and the algorithm returns to the first step when presented with another pattern. This perceptron is operant because it only learns when it chooses to act; if it does not choose to act on a given trial, its connection weights are not updated. Furthermore, it is operant in the way that Miller and Shettleworth (2007) desire, because as the associative strength of the cues at a location increase, P_L (a function of those associative strengths) increases, and so does the likelihood that location L will be visited. Conversely, as P_L decreases, so does the likelihood of visiting location L . As it learns about its environment, the operant perceptron will be more likely to choose locations (i.e., cue configurations) that lead to reinforcement, and will be less likely to choose locations that do not lead to reinforcement. As was noted earlier, and is detailed in the Technical Appendix, the root mathematical cause of the problems with the M–S model is the fact that it scales changes in associative weights by P_L . The procedure for training an operant perceptron solves this problem by separating operant choice from weight modification. That is, Equation 5 is used to make a decision about whether to learn or not, and then standard learning (Equation 6) is conducted accordingly. Importantly, at no point in Equation 6 are changes in association multiplied by Equation 5, preventing the mathematical difficulties described in the Appendix.

With this training algorithm, it is not clear how to relate a sweep of network training (one presentation of each location in a given environment) to a trial in an animal learning experiment. That is, it is unlikely that a network's sweep is identical to a subject's trial. However, such a direct comparison between a network and a biological agent is likely not useful, because one can modify the amount of learning accomplished by a network by manipulating the learning rate parameters, and such modifications are largely arbitrary. A more fruitful comparison to make involves what Pylyshyn (1984) has called relative complexity evidence. This evidence is obtained by comparing networks trained in one condition with networks trained in another, and then by relating this comparison to a similar one made by biological agents. For instance, if a network trained with three-feature objects learns the reorientation task in fewer epochs than a network trained with one-feature objects, and if a similar result was observed in an animal learning experiment, then this comparison would be meaningful, and would not depend on a strict quantitative mapping from network parameters to the animal learning paradigm.

Employing this algorithm, the operant perceptron can choose to investigate a location, or choose to not investigate a location; learning only occurs when the perceptron chooses to investigate. Furthermore, if one views the perceptron as a behaving agent, then the only behavior a perceptron would produce is when it chooses to investigate a location. As in geometry experiments with animals, when the agent (animal or perceptron) chooses to *not* undertake a particular course of action, this absence of behavior is naturally unobservable. If one removes the “unobservable” acts from the operant perceptron's history, there is no change to the perceptron's experience or behavior (as the perceptron did not

learn during these presentations), and what remains is a sequence of choice behavior that is more recognizable to animal researchers.

Previous research has shown that the operant perceptron can learn to perform a probability matching task (Dawson et al., 2009). Furthermore, this previous research has shown that the behavior of the operant perceptron at equilibrium is similar to that of a traditional perceptron trained on the same probability matching task. This indicates that the operant training procedure does not violate the mathematical regularities associated with perceptron learning. As we noted above, traditional perceptrons have been shown to be promising models of reorientation (Dawson et al., 2010). In the next section, we demonstrate that this is also true of the operant perceptron, by showing that it generates appropriate results for both of the case studies introduced earlier in this article.

Evaluating the Operant Perceptron

To see if the operant perceptron is capable of succeeding in which the M–S model had difficulty, we simulated both the reorientation and superconditioning tasks, as described below.

Reorientation

Our first simulation involved training the operant perceptron on the Wall et al. (2004) reorientation task. First, we modeled this task using Miller and Shettleworth's (2007, 2008) parameters, and then we tested for robustness by exploring the operant perceptron's behavior when different parameters were used.

Method. The Wall et al. task was illustrated above in Figure 1. The training set consists of four input units, used to represent the presence or absence of the B , F , G , and W cues. Any cue could be presented to the perceptron by activating its input unit with a value of 1; if a cue was absent, then the activity of its input unit was 0. B was initialized to have a 0.1 initial weight, and the other three cues had 0 initial weight. The network's learning rate was set to 0.15, exactly as in Miller and Shettleworth (2007). The training procedure was altered into an operant procedure as follows: After each pattern (i.e., collection of cues at a location) was presented to the network, the perceptron computed P_L using Equation 5, producing a number between 0 and 1. Then, a random number between 0 and 1 was generated. If the network's activity exceeded that random number, the network was said to have “chosen” to visit the location on this trial, and its connection weights were updated. If the network's activity did not exceed this number, the network was said to have not chosen to visit the location on this trial, and no weights were updated. This process repeats for every pattern (location) present in the scenario. Five such networks ran until convergence (after approximately 8,500 sweeps), and their responses were averaged. Each condition was modeled with this small number of repetitions because the variability between networks was low enough that more “subjects” were not required to establish the statistical significance of the observed effects.

Results. The results of this simulation are displayed in Figure 4. These results are virtually indistinguishable from the M–S 2008 model, if that model was supplied with well-behaved parameters. Specifically, the associative strengths of each cue asymptote (without fluctuation or singularity) at values that produce a high probability of investigating the Correct location, an intermediate (but low) probability of investigating the Rotational location, and a

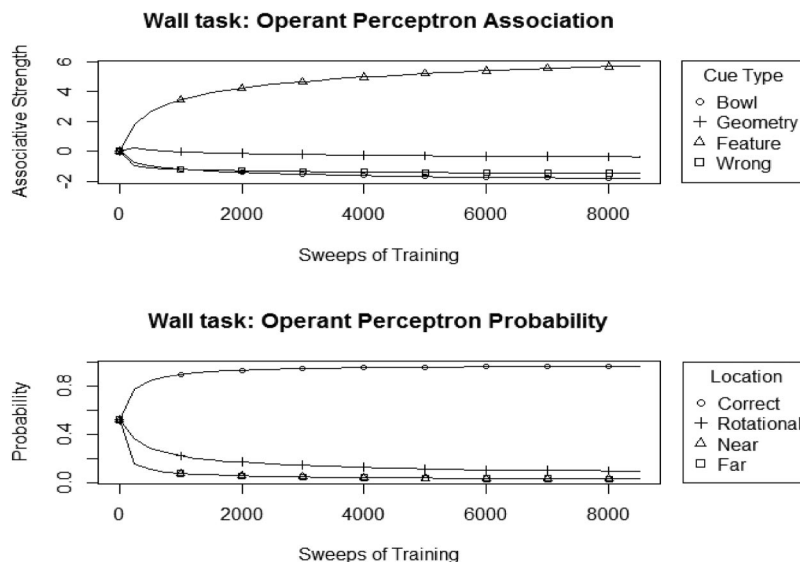


Figure 4. Top panel: Associative strengths acquired by the various cues (see Figure 1) during the operant perceptron's training on the Wall et al. (2004) reorientation task. Bottom panel: The probability with which the perceptron will choose to investigate the various locations during training. Each location is considered independent of the others, so these probabilities need not sum to 1 across all four locations.

near-zero probability of investigating the Near or Far location. It would seem that the operant perceptron model is capable of learning the reorientation task as defined by Wall et al. (2004).

In order to test for robustness, the operant perceptron model was run again with extreme learning rates under the same initial conditions. Setting the learning rate to 0.7 did not lead to fluctuations nor to any “crash”—rather, the perceptron converged normally after approximately 1,700 sweeps. Even a learning rate of 1.00—which causes the M–S 2008 model to collapse after just 12 iterations—results in the perceptron converging in $\sim 1,300$ sweeps. Continuing to simulate learning beyond convergence produces negligible changes from the pattern described here.

In conclusion, we find that this operant perceptron model is capable of empirically handling the reorientation results for which the Miller and Shettleworth (2007, 2008) models are acclaimed, and that the operant perceptron's behavior is robust to the extreme choices in learning rate that caused trouble for the M–S model.

Superconditioning

Our second simulation trained the operant perceptron on Horne and Pearce's (2010) geometric superconditioning experiment. We examined the perceptron's probability of responding to specific locations for evidence of superconditioning.

Method. Horne and Pearce's (2010) superconditioning experiment was presented to the operant perceptron using five inputs, corresponding to the five cues (B , G , W , F , and N) from Figure 3 above. These cues were grouped by location and stage as described in Figure 3, with each training pattern (i.e., each set of available cues) representing a corner present at a given stage. The network used a single output unit with a logistic activation function with bias held constant at 0. Based on the parameters described in

Horne and Pearce's experiment, the network's learning rate was set to 0.05, and all of its weights initialized to 0. Patterns were presented in a random order to the network on each pass through the training set. The perceptron's learning algorithm was made operant as described above.

As in Horne and Pearce (2010), these simulations ran for a specific, fixed number of “sweeps” before passing through to the next stage. Each stage received 5,000 sweeps of this training (sufficient to reach asymptote) before a final, geometry-only probe trial (consisting just of the $B + G$ cues) was carried out. Due to the stochastic nature of the operant perceptron's training procedure (which allows identical networks to make different patterns of choices), this simulation was repeated five times for each experimental condition and the aggregate responses were averaged. Each condition was modeled with this small number of repetitions because the variability between networks was low enough that more “subjects” were not required to establish the statistical significance of the observed effects.

Results. The operant perceptron's performance on this task over time during each of the four superconditioning stages is presented in Figure 5. As predicted above, the partially reinforced B and G cues acquire more negative associative strength during all stages than they did in the M–S 2008 model (reported in Horne & Pearce, 2010, Figure 11). After training, the experimental network had average connection weights (associative strengths) of $B = 1.3$ and $G = 4.91$, and the control network had connection weights of $B = 0.29$, $G = 3.20$. During geometry-only probe trials, the networks' mean responses to these cues—the logistic function of $B + G$ —are 0.998 for the experimental network and 0.970 for the control network.

Because these responses reflect the probability of investigating these correct-geometry-only locations, it can be concluded that both groups of networks were capable of learning the geometry of

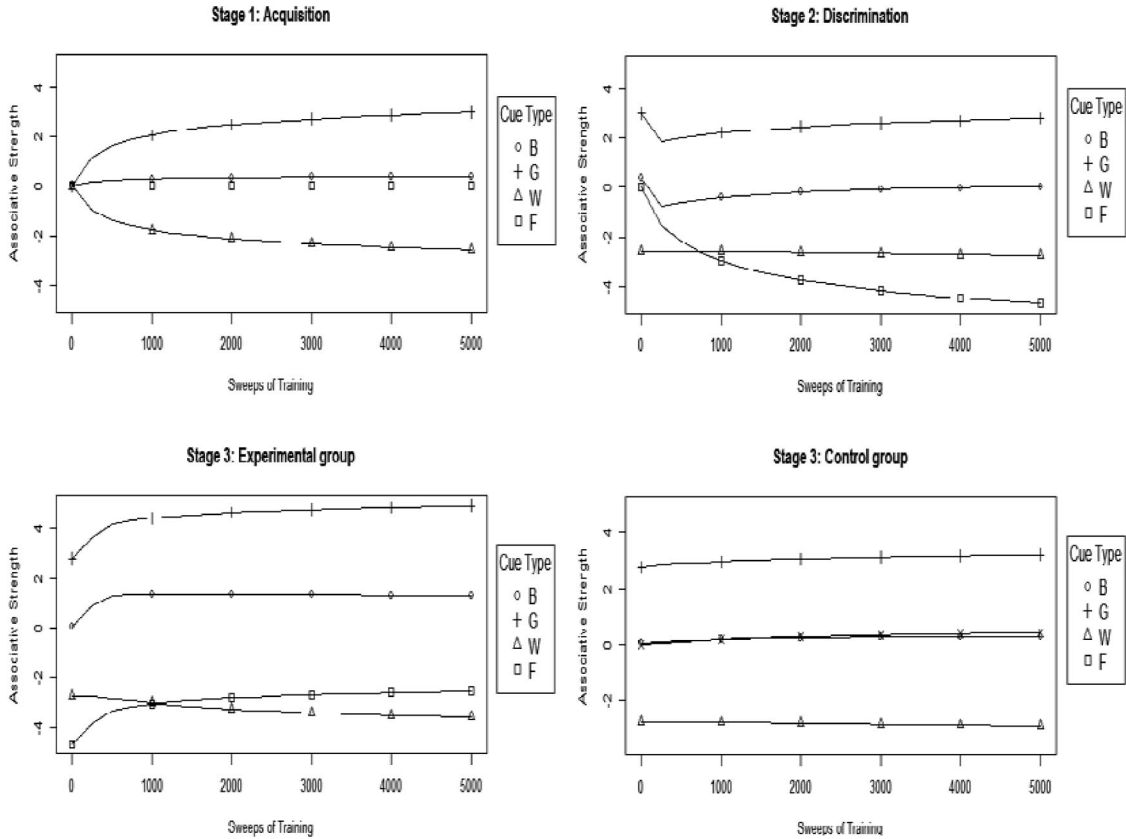


Figure 5. The operant perceptron learning the superconditioning task described above. The panels denote change in associative strength for each cue (B = “Base”, G = “Correct Geometry”, W = “Wrong Geometry”, F = “Reinforced Feature” and N = “Neutral Feature”) over time during each stage of this task. Compare with [Horne and Pearce \(2010\) Figure 11](#), which trained the M–S 2008 model on the same task but could not capture superconditioning behavior; the operant perceptron assigns several cues (B in particular) lower associative strength in comparison.

the task (they have a high probability of visiting a corner with only the correct geometry present). Additionally, the experimental group’s probability of investigating a geometrically correct corner is higher than that of the control group: The experimental group displays evidence of superconditioning. This is in agreement with [Horne and Pearce’s \(2010\) animal data](#), and distinct from their attempt to model the same task with the M–S 2008 model, which produced the opposite result (a response of 0.92 in the experimental condition, and a response of 0.95 in the control condition). This difference is small (indeed, almost identical to the difference seen in [Horne and Pearce \(2010\)](#)), but statistically significant, $t(4) = -63.11$, $p < .01$, and the difference is in the correct direction to conclude that the operant perceptron succeeds on modeling superconditioning, in which the M–S 2008 model did not. Comparing the operant perceptron’s learning (see [Figure 5](#)) with the Miller–Shettleworth model’s behavior ([Horne & Pearce, 2010, Figure 11](#)) makes it clear that the M–S 2008 model fails to assign sufficient inhibitory associative strength to several cues (B and G in particular) in Stage 2 and both conditions of Stage 3, because acquisition of inhibitory associative strength is scaled down in the Miller–Shettleworth model as described above.

General Discussion

[Miller and Shettleworth’s \(2007, 2008\) associative model of reorientation](#) is rooted in the observation that reorientation is fundamentally a problem of operant learning. Therefore, they endeavored to formalize this by creating an operant version of the established [Rescorla–Wagner \(1972\) theory of associative learning](#). However, their attempt to extend the Rescorla–Wagner model in this fashion had problems.

Some of these problems have already been documented in the literature. [Dawson et al. \(2008\)](#) discovered that the original M–S model (2007) will produce impossible probabilities under a variety of circumstances. This led Miller and Shettleworth to modify their original model ([Miller & Shettleworth, 2008](#)). However, [Horne and Pearce \(2010\)](#) found that this revised model did not correctly model animal data collected for tasks to which the model should apply, such as geometric superconditioning.

Other problems with the M–S model have been reported for the first time in the current article. As shown above, when model parameters are manipulated, the behavior of the M–S model is unstable, producing dramatic fluctuations. Indeed, for the M–S

2008 model, these fluctuations lead to an eventual “crash” that is caused when the model is required to divide a value by zero.

Although previous studies (Dawson et al., 2008; Horne & Pearce, 2010) have documented some problems with the M–S models, they did not locate the source of these difficulties. In addition to discussing some new problems, we have also shown that all of these problems emerge from the structure of Equation 2. In effect, the equation cannot properly handle situations with uniquely inhibitory cues when those cues are paired with other, excitatory cues. These situations lead to inappropriate scaling of changes in associative strength when Equation 2 is employed, and this ultimately gives rise to all of the problems described above.

However, in the current article we have done more than demonstrate problems and trace their mathematical root. We have also provided a different model that can overcome these problems. Our alternative formalism preserves Miller and Shettleworth’s (2007) operant-learning goal, but is anchored on the solid foundations of artificial neural network mathematics (e.g., Rosenblatt, 1958, 1962). We have presented simulations that show this new “operant perceptron” model corrects these problems, and is capable of learning both reorientation and superconditioning. We believe this operant perceptron to be a plausible architecture for modeling reorientation task learning, which was the primary intent of the M–S model (Miller, 2009; Miller & Shettleworth, 2007, 2008). We have already shown that the operant perceptron can easily model other domains; for instance, it can learn to match reinforcement probabilities (Dawson et al., 2009). The extent to which the operant perceptron can match the ability of the traditional perceptron to model the further intricacies of reorientation (Dawson et al., 2010) or to model a variety of classical conditioning paradigms (Dawson, 2008) is clearly a matter for future research.

References

- Cheng, K. (1986). A purely geometric module in the rat’s spatial representation. *Cognition*, 23, 149–178. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/3742991> doi:10.1016/0010-0277(86)90041-7
- Cheng, K., & Newcombe, N. S. (2005). Is there a geometric module for spatial orientation? Squaring theory and evidence. *Psychonomic Bulletin & Review*, 12(1), 1–23. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/15945200> doi:10.3758/BF03196346
- Chiandetti, C., & Vallortigara, G. (2008). Spatial reorientation in large and small enclosures: Comparative and developmental perspectives. *Cognitive Processing*, 9, 229–238. doi:10.1007/s10339-008-0202-6
- Cramer, J. S. (2003). *Logit models from economics and other fields*. Cambridge, MA: Cambridge University Press. doi:10.1017/CBO9780511615412
- Dawson, M. R. W. (2004). *Minds and machines: Connectionism and psychological modeling*. Oxford, UK: Blackwell. doi:10.1002/9780470752999
- Dawson, M. R. W. (2008). Connectionism and classical conditioning. Comparative cognition and behaviour reviews. *Comparative Cognition Society*, 3, 115. doi:10.3819/ccbr.2008.30008
- Dawson, M. R. W., & Dupuis, B. (2012). The equilibria of perceptrons for simple contingency problems. *IEEE Transactions in Neural Networks and Learning Systems*, 23, 1340–1344. doi:10.1109/TNNLS.2012.2199766
- Dawson, M. R. W., Dupuis, B., Spetch, M. L., & Kelly, D. M. (2009). Simple artificial neural networks that match probability and exploit and explore when confronting a multiarmed bandit. *IEEE Transactions in Neural Networks*, 20, 1368–1371. doi:10.1109/TNN.2009.2025588
- Dawson, M. R. W., Kelly, D. M., Spetch, M. L., & Dupuis, B. (2008). Learning about environmental geometry: A flaw in Miller and Shettleworth’s (2007) operant model. *Journal of Experimental Psychology: Animal Behavior Processes*, 34, 415–418. doi:10.1037/0097-7403.34.3.415
- Dawson, M. R. W., Kelly, D. M., Spetch, M. L., & Dupuis, B. (2010). Using perceptrons to explore the reorientation task. *Cognition*, 114, 207–226. doi:10.1016/j.cognition.2009.09.006
- Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, 117, 227–247. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/2971760> doi:10.1037/0096-3445.117.3.227
- Horne, M. R., & Pearce, J. M. (2010). Conditioned inhibition and superconditioning in an environment with a distinctive shape. *Journal of Experimental Psychology: Animal Behavior Processes*, 36, 381–394. doi:10.1037/a0017837
- Miller, N. (2009). Modeling the effects of enclosure size on geometry learning. *Behavioural Processes*, 80, 306–313. doi:10.1016/j.beproc.2008.12.011
- Miller, N. Y., & Shettleworth, S. J. (2007). Learning about environmental geometry: An associative model. *Journal of Experimental Psychology: Animal Behavior Processes*, 33, 191–212. doi:10.1037/0097-7403.33.3.191
- Miller, N. Y., & Shettleworth, S. J. (2008). An associative model of geometry learning: A modified choice rule. *Journal of Experimental Psychology: Animal Behavior Processes*, 34, 419–422. doi:10.1037/0097-7403.34.3.419
- Pylyshyn, Z. (1984). *Computation and cognition*. Cambridge, MA: MIT Press.
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York, NY: Appleton-Century-Crofts.
- Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65, 386–408. doi:10.1037/h0042519
- Rosenblatt, F. (1962). *Principles of neurodynamics*. Washington, DC: Spartan Books.
- Sutton, R. S., & Barto, A. G. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review*, 88, 135–170. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/7291377>
- Wall, P. L., Botly, L. C. P., Black, C. K., & Shettleworth, S. J. (2004). The geometric module in the rat: Independence of shape and feature learning in a food finding task. *Learning & Behavior*, 32, 289–298. doi:10.3758/BF03196028

(Appendix follows)

Appendix

Technical Appendix

In the prior sections, we described a number of problems in the Miller and Shettleworth (2007, 2008) associative models. We described these problems purely in terms of the behavior of the models. However, the source for this erroneous behavior is ultimately mathematical, and emerges primarily from Miller and Shettleworth's choice to scale the Rescorla–Wagner (1972) equation by Equation 3 or Equation 4. This Appendix will elaborate upon this, demonstrating why such scaling is incorrect from the perspective of calculus, and why such scaling produces incorrect results.

The main characteristic of the Miller and Shettleworth (2007, 2008) model is that it multiplies the Rescorla–Wagner (1972) equation by a probability term. The intention of this multiplication is to make the Rescorla–Wagner model operant in nature. However, it is this multiplication that causes the problems that were identified earlier in this article. This multiplication also causes the model to make unexpected (and, we believe, unintentional) claims about time.

In Rescorla and Wagner (1972), the parameter β is explicitly defined as a learning rate parameter—a rate reflecting how much learning takes place within a given amount of time. β is held constant when the model is employed, because to do otherwise would “beg justification” (p. 82). A consequence of holding learning rate constant over each iteration of the model is that the model implies a constant amount of time passes during each iteration. However, because the learning rate β is held constant, this “time step” is usually suppressed when writing the equations.

We now express the Rescorla–Wagner equation in terms consistent with Miller and Shettleworth's (2007) approach, making time explicit:

$$\Delta V_i(t) = \alpha \beta (\lambda - \sum V_i(t_i)) \quad (\text{a1})$$

Here, the subscript i refers to iteration: the change in weights from the current iteration to the next depends on the sum of the weights at the current time. In this equation, the learning rate parameter β is proportional to the amount of time that passes per iteration, $(\Delta t)/(\Delta i)$. If the equation is consistently applied to every cue at every iteration, then $\Delta i = 1$, and thus β is proportional to Δt . Because Miller and Shettleworth always apply their equation at every iteration, this assumption holds. Therefore, by subsuming the constant of proportionality into α and setting $\beta = \Delta t$, we can substitute into Equation a1 and rearrange to form a ratio, as follows:

$$\Delta V_i(t) = \alpha (\lambda - \sum V_i(t_i)) \Delta t \rightarrow \frac{\Delta V_i(t)}{\Delta t} = \alpha (\lambda - \sum V_i(t_i)) \quad (\text{a2})$$

By the definition of a limit, as the time step Δt approaches zero, this ratio will approach the instantaneous time derivative of associative strength, and thus:

$$\lim_{\Delta t \rightarrow 0} \frac{\Delta V_i(t)}{\Delta t} = \frac{\partial V_i(t)}{\partial t} \quad (\text{a3})$$

Because this equivalence holds, the model's associative strength will change at the same rate on every iteration (Equation a2) as it does at each unit of time (Equation a3).

In defining their models, Miller and Shettleworth (2007, 2008) multiply the entire Rescorla–Wagner equation by some probability term P_L . Both the 2007 (Equation 3) and 2008 (Equation 4) P_L terms are functions of associative strength at a particular iteration V_i , which is itself a function of time, and therefore Equation 3 and 4 are both also functions of time. Both P_L equations can be expressed more generally as:

$$P_i(V_i(t)) = (P \circ V)(t_i) \quad (\text{a4})$$

Simply multiplying this term into the Rescorla–Wagner equation, as Miller and Shettleworth (2007, 2008) have done in Equation 2, introduces a second time dependency to the system:

$$\alpha (\lambda - \sum V_i(t_i)) * (P \circ V)(t_i) = \frac{\Delta V_i(t)}{\Delta t} * (P \circ V)(t_i) \quad (\text{a5})$$

If $\beta = 1$ (and, as a consequence, $\Delta t = 1$), this equation reduces to the right side of Equation 2, which Miller and Shettleworth label “ ΔV ”. However, this is *not* the equivalent of ΔV if $\Delta t = 1$, rendering Equation a3 invalid:

$$\lim_{\Delta t \rightarrow 0} \left[\frac{\Delta V_i(t)}{\Delta t} * (P \circ V)(t) \right] \neq \frac{\partial V_i(t)}{\partial t} \quad (\text{a6})$$

Rather, the proper time derivative would handle this second time dependency through applying the chain rule to Equation a4 and accounting for the resulting $\Delta V/\Delta t$ term (the exact form of which would depend on whether Equation 3 or Equation 4 was used for P_L). Miller and Shettleworth (2007, 2008) did not do this, but continued to treat the composite Equation 2 as if it were supplying a proper change over time. Instead, the additional uncontrolled time dependency causes time in the simulation to flow at different rates for each location, depending on the current net associative strength of that location.

(Appendix continues)

This gives rise to the scaling problems discussed above: The scaling introduced by different P_L values is effectively changing the learning rate at each location on a given iteration, without the justification that Rescorla and Wagner (1972) “beg,” which results in different amounts of time passing between iterations at each location (including exactly zero time between iterations for net-negative locations in the M–S 2008 model, as $P_L = 0$ in those cases). However, the model determines this scale by referencing weights at a given *iteration*, instead of after a given amount of *time* has passed—and due to the effective learning-rate scaling introduced by P_L , these are no longer equivalent considerations (Equation a6). In effect, this scaling gives rise to a situation in which the change in weight at one iteration may alter the next moment for one cue, but some moment in the past for a different cue, allowing the future to influence the past. This is, of course, impossible.

For illustration, we return to the Wall et al. (2004) reorientation task described earlier, using Miller and Shettleworth’s (2007, 2008) original parameters. Figure A1 directly compares both the M–S 2008 and M–S 2007 models’ behavior reported in terms of iteration with their behavior reported in terms of time. Normally for Rescorla–Wagner models, the amount of time that passes at each iteration is defined by β , and scaled by the dimensionless salience term α ; time passing per iteration is therefore proportional to $\alpha\beta$ (a constant). With Miller and Shettleworth’s Equation 2, these are further scaled by the dimensionless P_L term, which varies over time. Therefore, in Figure A1, the “time” axis reflects the cumulative value of $\alpha\beta P_L$ at each iteration.

Observe from Figure A1 that both models consider “simultaneous events” (such as the highlighted 60th iteration) which actually reflect different points in time. The nature of this asynchrony depends on the choice of equation for P_L . If Equation 3 is used, P_L

is allowed to go negative for locations with sufficiently inhibitory net associative strength (Dawson, Kelly, Spetch, & Dupuis, 2008), therefore Δt becomes negative and time begins to flow backward. The singularities in the M–S 2007 model form immediately following when $t < 0$ for the Wrong cue. (This corresponds to the exact point in which ΣV_L changes signs in Equation 3.) The positiveness correction employed by the M–S 2008 model prevents time from flowing backward, but does not prevent a given iteration from reflecting different points in time. In fact, when Equation 4 sets P_L to 0, Δt is also set to 0 for the corresponding cue (Wrong), such that every subsequent iteration reflects the same point in time for locations containing that cue.

Behavior of this nature will show up as a natural consequence of the mathematical choice to multiply Equation 1 by some other function (creating an alternate form of Equation 2), regardless of the nature of that other function, without also controlling for the uncontrolled time dependency. Simply replacing Equations 3 and 4 with Equation 5 (the logistic function) to produce a “logistic M–S model,” for instance, will still produce errant behavior, as Equation a6 still holds in such a case.

In contrast, the operant perceptron model presented above does not suffer from this problem, as the underlying mathematics for updating its connection weights are formally equivalent to the Rescorla–Wagner equation (Dawson, 2008), and no scaling is applied. Although the exact sequence of locations visited by the network may vary the amount of time the network spends at each location, this is functionally equivalent to adjusting the number of times each location is presented to the network—a course of action that does not introduce any uncontrolled time dependencies into the calculations for ΔV .

(Appendix continues)

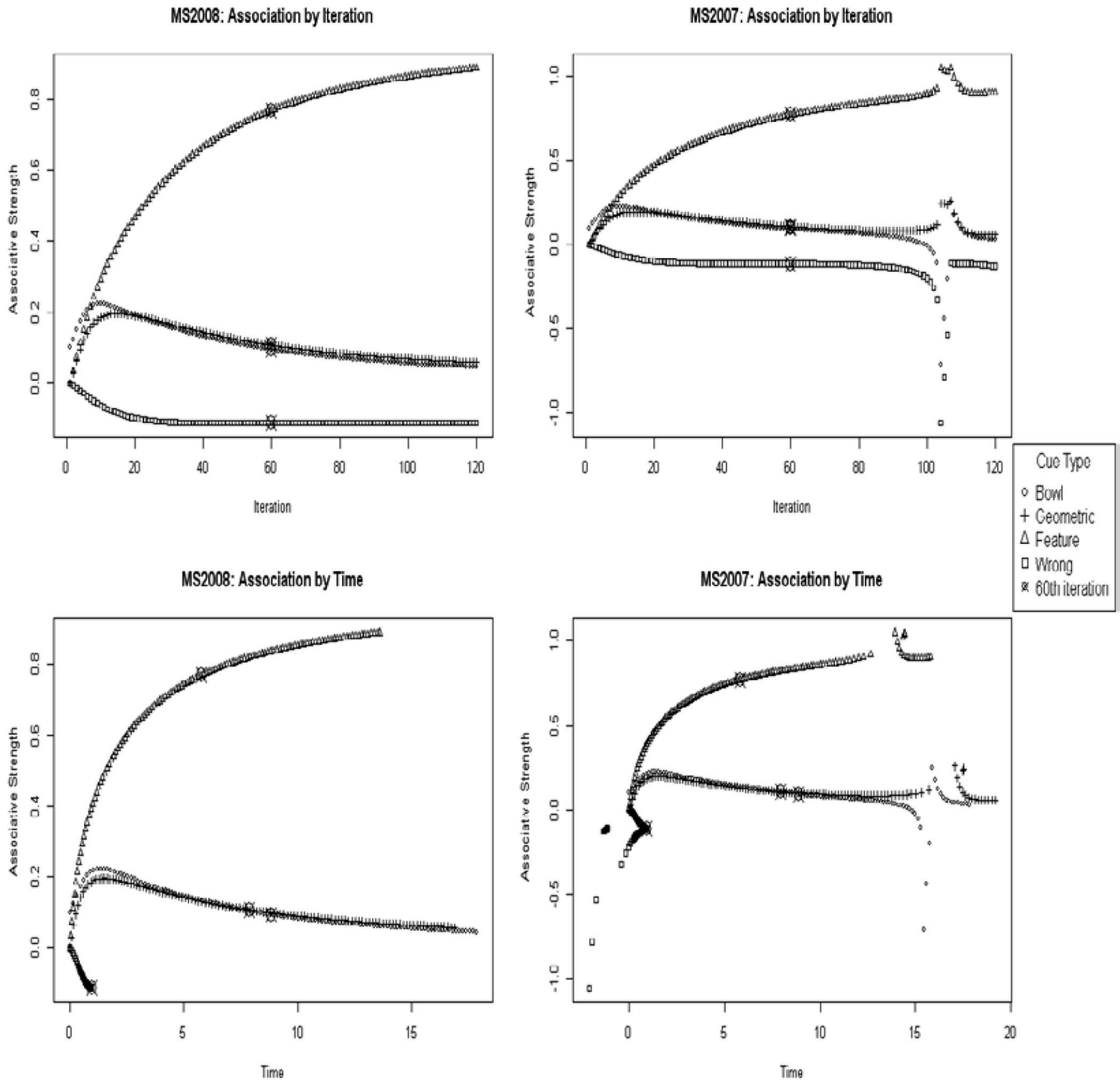


Figure A1. The Wall et al. (2004) task, as interpreted by both the M-S 2008 (left) and M-S 2007 (right) models. The lower panels rescale the horizontal axis by $\alpha\beta P_L$, becoming proportional to time that has passed for each cue; the 60th iterations (considered to be simultaneous by the model) are highlighted.

Received July 17, 2012
 Revision received January 22, 2013
 Accepted February 4, 2013 ■