Rethinking the role of error in attentional learning

Mark R. Blair (mark.blair@sfu.ca)
R . Calen Walshe (calen.walshe@sfu.ca)
Jordan I. Barnes (jordanb@sfu.ca)
Lihan Chen (bill.lihan@gmail.com)

Department of Psychology, Cognitive Science Program
Simon Fraser University, 8888 University Drive
Burnaby, BC, V5A 1S6 Canada

Abstract

Learning how to allocate attention properly is essential for success at many tasks. Extant theories of categorization assume that learning to allocate attention is an error-driven process, where shifts in attention are made to reduce error. The present work introduces a new measure, error bias, which compares the amount of attentional change in response to incorrect responses versus correct responses during category learning. We first confirm that prominent categorization models predict high amounts of error bias. We then test this prediction against human eye-tracking data from 384 participants. Across 7 of 8 data sets we find that participants show minimal or no error bias. This finding suggests that attentional learning mechanisms, as implemented in influential computational models, cannot be generalized to account for measures of overt attention.

Keywords: Attention; Error; Eyetracking; Categorization; Eye-Movements; Optimization; Learning; Modeling

Introduction

Giraffes have long necks, helicopters have propellers on top, and wedding cakes are taller than birthday cakes. Learning these categories often involves learning to attend to such highly predictive features. This kind of selective use of information is present very early in human development. For example, infants focus mostly on the head to discriminate cats from dogs (Quinn, Doran, Reiss, & Hoffman, 2009), but they use legs and wheels when distinguishing animals from vehicles (Rakison & Butterworth, 1998). People also learn to change how they attend to stimuli with experience, and experts with years of training can develop the ability to use subtle but highly informative stimulus dimensions (Biederman & Shiffrar, 1987). Although the process by which people learn the right information to attend – what we shall call attentional learning – is a critical part of learning, from nascent stages to the highest levels of performance, its mechanisms are not well understood. Though overt attentional allocation can be studied directly and relatively accurately with modern eye-tracking, there is no existing theory that makes specific behavioural predictions about how the allocation of overt attention changes during learning. Our work is intended to be some early steps toward the goal of building such a theory.

Although there is not an existing theory intended to account for attention at the level of eye-movements, the literature on category learning has theories which contain precise descriptions of attentional learning more generally (e.g., Kruschke, 1992). Researchers have created formal, computational models of how the effective allocation of attention is learned, and how it interacts with perception, memory and decision-making to improve categorization performance. In these computational theories, attention is characterized as a weight on each stimulus feature that is adjusted to reduce error, and more specifically, adjusted such that the proportion of change is relative to the proportion of error.

In the present study, we use eye movements as an index of attention and compare those measures to the model equivalents. It is not always clear what attentional weights in models are supposed to correspond to in the real world. Attention is a complex series of independent means of biasing information processing. One very important source of such biases is the overt manipulation of sensory receptors, like eye movements and although these models were not intended to account for eye-movements directly, given that there are tight connections between covert and overt forms of attention, they are an excellent starting place (McPeek, Maljkovic, & Nakayama 1999). Indeed, it is clear from several recent eye-tracking studies (Blair, Watson, Walshe, & Maj, 2009; Rehder & Hoffman, 2005a, 2005b) that over the course of an experiment people get better at ignoring irrelevant information as they get better at categorizing, a finding in accord with existing error-driven accounts.

There is some evidence, though, that error may not be the sole ingredient for attentional learning. Bott, Hoffman and Murphy (2007) have shown that participants attend to more dimensions than are strictly necessary to perform well. Blair, Watson, and Meier (2009) have shown that participants continue to optimize their attention, even after feedback is removed and participants have stopped making errors. Rehder and Hoffman (2005a, 2005b; Kim & Rehder, 2009) have found that reductions in the probability of fixating irrelevant information occur several trials after reductions in incorrect responding, rather than before as one might expect. While these studies are suggestive, they are not necessarily a requiem for error-driven accounts.

In the formal theories discussed above, the error that motivates change is error internal to the model, not response error. The difference is subtle, but important. Imagine believing that hockey team A was only slightly better than team B. You might reevaluate your estimates of team A's skill if they won ten games in a row, even though you would predict them to win every time. Your estimate of the teams' relative strengths was erroneous, even though your predictions were correct. Internal error is calculated in a similar way in the models; attention is adjusted based on mismatches between feedback and the model's internal estimates, not on the model's actual choices. Because of this fact, it is possible for error-driven models to predict some attentional shifting even without incorrect responses as long as the internal estimates do not completely match the feedback.

Despite the indirect connection between response errors and attentional shifts, error-driven models still make testable predictions about their relationship. The larger the model's internal error, the more likely it is to make a performance error, and internal error should, on average, be higher on incorrect trials. Thus, error-driven models of attentional change might be said to predict an error bias. This prediction is straightforward to confirm by running simulations of the model, and, unlike predictions about internal error, is also easy to test in humans. While it may seem as though an error-bias would be a simple property of the model to predict solely based on the mathematics of back-propagation, this is not actually the case. Complex models like ALCOVE and RASHNL contain many explicit and structural parameter settings that have non-linear interactions with one another. These parameters could have a number of attenuating influences on the degree of error-bias and the only real test of these interactions is simulation.

The goal of the present work is to empirically evaluate the idea that attentional allocation is adjusted in proportion to error during learning. In our study, we use a novel measure, the error bias, that is calculable from both formal models of category learning and human eye-gaze data. We first run simulations of two popular models of categorization, ALCOVE and RASHNL, to confirm that they indeed predict a strong error bias. We then compare the error bias distributions from our simulations to 8 sets of human data gathered from a variety of category structures.

Recent thought in the methodology of model evaluation holds that a model should be evaluated on a broader basis than simply its performance with best-fitting parameters. For instance, Roberts & Pashler (2000) argued that a model can be considered a good fit to the data if it matches both the central tendency and the variability of the data. Following that idea, if the model assumptions (error-driven learning, proportional adjustment) are correct, we should see not only similar mean error bias scores, but similar distributions of error bias scores between the models and the human data.

Model Simulations

We chose to use the category learning models ALCOVE (Kruschke, 1992) and RASHNL (Kruschke & Johansen, 1998) as prototypical error-driven attentional learning models. They are both popular models that enjoy widespread usage as benchmarks for current models (Little & Lewandowsky, 2009; Love, Medin, & Gureckis, 2004) Using them as a basis for our analysis allows our results to generalize widely. Both models are similar in that they rely on gradient descent on error to minimize the distance between the predicted and observed values on the output layers. In addition, both models have an attentional layer that modulates the gain of the features presented to the input layer. However, RASHNL extends ALCOVE in two important ways: first, attention shift is iterated multiple times on a single training instance, resulting in faster overall attention shifting. Second, RASHNL incorporates annealed learning whereby the learning rate is reduced as a function of trial number, allowing the model to settle into a consistent response pattern despite early fast attention shifting.

Method

The error bias measure. The models shift attention by changing the attentional weights assigned to stimulus dimensions. We refer to the amount by which attentional allocations change between trials as the attentional change. If the allocation of attention is the same to all stimulus dimensions for the current trial as it was for the previous trial, attentional change would be 0. If, on the other hand, the model goes from attending only to dimension 1 (i.e., a weight of 1 on that dimension and a weight of 0 on the other stimulus dimensions) to attending only to dimension 2 (i.e., a weight of 1 on dimension 2 and weight of 0 on the other dimensions) then the attentional change will be 2. Our measure of error bias is the difference between the mean attentional change following error trials and the mean

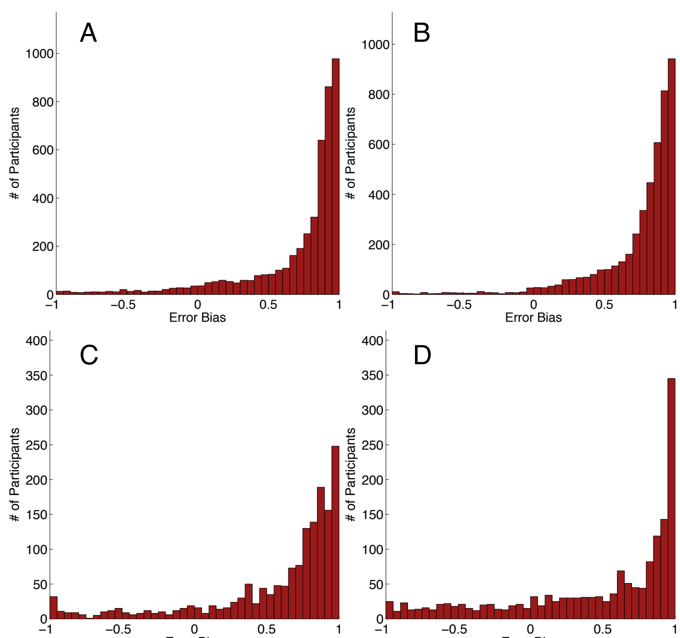

Figure 1. Simulated Distributions of Error Biases. A. ALCOVE model: randomly-sampled parameter values that met the learning criterion. B. RASHNL model: randomly-sampled parameter values that met the learning criterion. C. ALCOVE model: best-fitting parameter values. D. RASHNL model: best-fitting parameter values. See text for details.

attentional change following correct trials over the total of the two. The error bias is +1 in cases where all the attentional change occurs following incorrect trials, 0 when there is no difference in the amount of attentional change following incorrect and correct trials, and -1 when all attentional change occurs following correct trials.

The random sampling procedure. We conducted a number of simulations to confirm the prediction that models which use an error signal to drive attentional shifting will exhibit an error bias. While an error bias can be predicted from the equations underlying of the model, the simulations allow us to translate more directly to experimental results. Our primary concern was the extent to which a random sampling of the parameter space would result in models that consistently predict a large error bias. The sampling for ALCOVE was conducted along the following dimensions and ranges: specificity (.01 - 40), choice decisiveness (1 - 10), attention shift rate (0 - 50) and output learning rate (0 - 3). In extending the analysis to RASHNL, sampling was also conducted on the annealing parameter (0 - 1). To add fast attention shifting to the model the number of attentional shift iterations set to 10. These bounds contain the historically best-fitting parameters for these models (Kruschke, 1992; Kruschke & Johansen, 1998).

Simulated experiment. We taught the models to classify four categories of stimuli that had three binary-valued dimensions. Categories are deterministic, and are a function of the four combinations of two of the dimensions. The value of the third dimension was equally indicative of all categories, and therefore non-diagnostic. On each trial, the model was given the values of the three dimensions for the presented stimulus. The model then produced a response which was a weighted random selection based on the response probabilities produced by the model. After that, the model was given feedback to use as a teaching signal to make adjustments to its memory and attentional components. The experiment was 360 trials long. We calculated the error bias based on the changes in the models' attentional weights that occurred directly after correct and incorrect trials. Stimulus presentation order was identical to the trial orderings generated for our human subjects. For both models, each participant's presentation order was used in conjunction with 100 randomly selected sets of parameter values to generate a distribution of error bias values.

Results

It became immediately clear that large regions of the parameter space produce models which perform erratically: either not learning at all, or crashing the simulation when variables grow to infinity. To restrict our exploration to reasonable parameter settings we excluded any parameter values under which the model did not meet a learning criterion of 9 consecutive correct trials, which was the same criterion that human participants were expected to meet in the original experiment. The resulting predictions of ALCOVE and RASHNL can be seen in Figures 1A (M=.70, Mdn=.86) and B (M=.74, Mdn=.86), respectively. In both figures the modal prediction is the maximal error bias of 1. Overall, the models are most likely to produce high bias values. Nevertheless, both models predict a maximally wide distribution of error-bias scores. This is a positive attribute of the model only if humans also produce a broad distribution of scores.

Human Data Fits

One possible limitation to the above simulations might be that many of the thousands of random model samples lead to response patterns that are nothing like human performance. As a result, the general model predictions may not be representative of the predictions from parameter settings which better characterize human performance in this particular task. To investigate this possibility, we fit both models directly to the human response data for this experiment. Using a bounded simplex search with a maximum of 5000 iterations, we found the 100 best-fitting parameter values for each of the 16 human subjects.

The error bias values produced by these best-fitting parameters are shown in Figure 1, C and D. The ALCOVE distribution is shown in panel C (M=.57, Mdn=.76) and the RASHNL distribution is shown in panel D (M=.44, Mdn=.67). Using only the best-fitting values removed many of the maximal error biases, but the histograms share the primary properties of the results from random sampling: high error biases are far more likely than low error biases, and the models produce values across the full range possibilities.

Having produced specific quantitative predictions for the distribution of error biases in this task, we compare the simulation results to human data.

Human Subjects Data

In this section we analyze the data from a number of eye-tracking experiments to assess the error-bias under a variety of tasks and conditions. These range from simple unidimensional rules to complex categories where different features are relevant for different categories. For readability, these category structures are introduced as we discuss the error-bias distributions from those experiments. We compare this to the high median and broad range of the error biases predicted by the error-driven models.

Method

The human data reported here were collected using a standard category learning paradigm. In these tasks participants are asked to learn to classify stimuli based on the values of prominent features. A trial consists of the presentation of the stimulus, the participants' category choice, and feedback specifying the correct category. In the current experiments, as well as in the paradigm generally, learners tend to improve both in answering correctly and in allocating less attention to irrelevant features. Unless otherwise noted, all data sets reported below used categories with two relevant and one irrelevant dimension. If the presented category structure comes from a previously published study, this is indicated. In all studies eye movement data were recorded using an eye-tracker sampling at 50, 60, or 120Hz (always consistent within an experiment) with a spatial resolution of 0.5°. Fixations were defined using a modified dispersion threshold algorithm with thresholds at 1° and 75ms. Eye movements were counted as fixations to category features if they fell within 100 pixels (≈ 1.9°) of a feature's centre. Shifts of attention were calculated in the same way for the humans as for the models, except that the proportion of time spent fixating the three stimulus features constituted the raw attention data on each trial, instead of the model's attention weights.

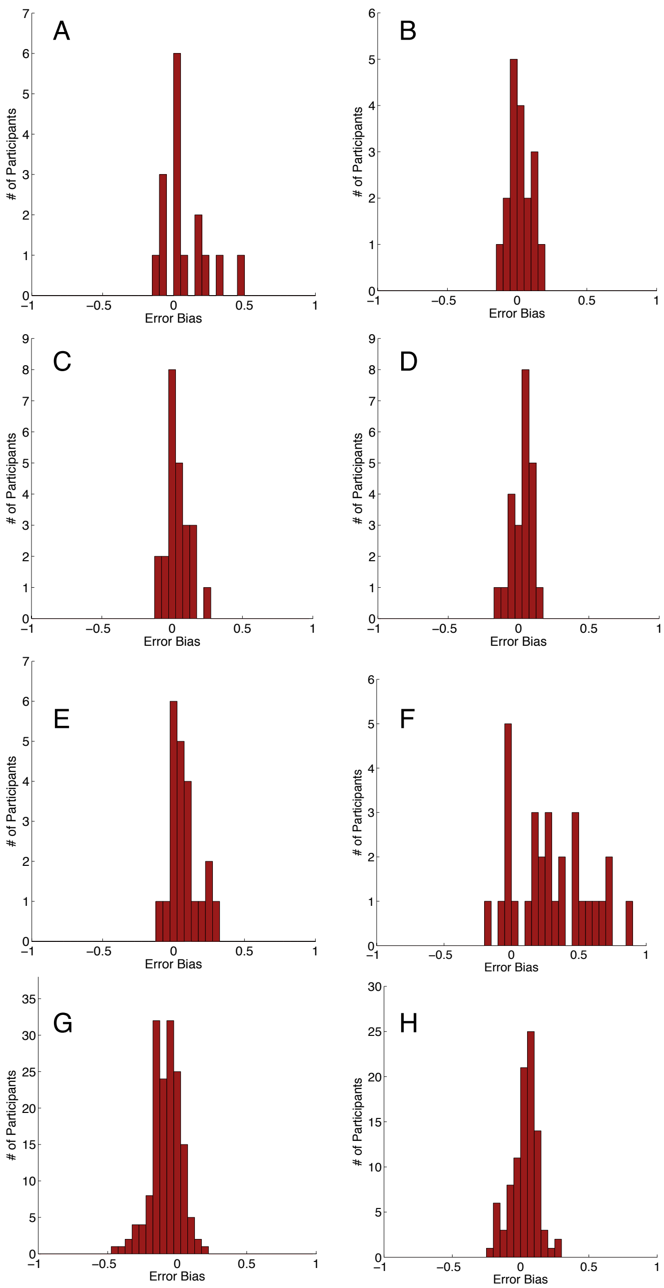

Figure 2: Human error bias distributions from eight different data sets. See text for descriptions.

Results and Discussion

The distribution of human error biases are plotted in Figure 2. Each panel represents a unique category structure. Panel A shows the human data from the experiment used in the model simulations reported above (M=.08, Mdn=.03). There is a stark contrast between the distribution of error bias scores of the human subjects and the model predictions shown in Figure 1. Two sample t-tests (unequal variances assumed) confirm that the human data are significantly different than all four of the simulations (ALCOVE, randomly sampled parameters - $t(29.73)=7.99$, $p<.001$; RASHNL, randomly sampled parameters - $t(29.46)=8.91$, $p<.001$; ALCOVE, best-fitting parameters - $t(32.54)=5.35$, $p<.001$; RASHNL, best-fitting parameters - $t(33.97)=3.03$, $p<.01$). The remaining panels in Figure 2 show human error biases from a number of other categorization experiments performed in our laboratory.

Panel B shows the error biases of 18 participants learning a two-category task with continuous dimensioned stimuli, very similar to Blair and Homa (2005, Experiment 2). The mean of this distribution is .02, the median .04. Panel C plots the error biases of 23 participants who learned a four-category information integration category structure akin to that in Maddox, Filoteo, Hejl, & Ing (2004). The mean of this distribution is .03, and the median is .05. Panel D plots the error bias of 23 participants in a study using a four category rule-based structure (again, similar to Maddox et al., 2004). The mean of this distribution is 0.04, the median = .02. Panel E shows data from a two-category information integration task with 24 participants (M=.08, Mdn=.04). This kind of category has seen wide use in the category learning literature (e.g., Ashby and Gott, 1988).

Figure 2, panel F shows the first data set to depart from the very consistent pattern found in the other data sets (M=.29, Mdn=.25). The 30 participants learned a two category single-dimensional rule-based category structure similar to that used by Maddox and Ashby (2004). As can be seen, this data set shows a pronounced bias toward shifting on error trials (Mdn=0.26, M=0.30, $t(29)=2.61$, $p<.05$). This is also the only data set for which the categories had only one relevant feature. We discovered, upon looking at the error and attentional allocation data of the most biased subjects, that this effect seems to result from an extraordinary stability in the gaze data once the categories have been learned. Participants in this task are much better at restricting their gaze to relevant data, and so there is very little change on the correct trials that occur once the rule has been learned. If this explanation is correct, we should expect that category structures for which the optimal allocation of attention is difficult may show less error bias. The next data set is from such a task.

Panel G in Figure 2 plots data from the stimulus-responsive attention structure used in Blair, Watson, Walshe, et al., (2009, Exp 2). Each category has two relevant and one irrelevant dimension, but unlike the previous category structures used, the relevant features are not the same for every category. Features 1 and 2 are relevant for two of the four categories, and features 1 and 3 are relevant for the other two; participants can optimize their attention by looking at either 2 or 3, based on the value of feature 1.

Attentional optimization is more difficult in this task than in the other data sets and participants continue to slowly optimize attention well after performance errors have ceased. This has the effect of shifting the error bias into the negative values for most of the 156 participants (M=-.08, Mdn = -.08; one-sample t-test, $t(155)=107.91$, $p<.001$).

The final data set (Figure 2, panel H) consists of data from 95 subjects presented with the same category structure as the previous data set. This study was designed to investigate asymptotic attentional optimization in this complex task, and thus extended the original study (shown in panel G) by an average of 280 trials. With this extended training, participants in this study eventually stabilize the relative allocation to the various stimulus features. The small negative error bias that occurs in panel G has disappeared and the mean error bias is .02, and the median is .03, very similar to the previous data sets. This supports the idea that the error bias measure is sensitive to the difficulty of learning the optimal attentional allocation. Interestingly, the data from panel H are collapsed across two between-subjects conditions. One condition exactly matched that of the previous data set (panel G), except for the training duration; the other condition is one where one pair of categories was 5 times more common than the other. While this frequency manipulation dramatically changes the pattern of attentional allocation to particular features such that Feature 2 is often fixated before Feature 1, there is no discernible difference in the error bias of these conditions. Overall, human error biases seem highly consistent across a variety of structures and manipulations.

General Discussion

The present work investigated the extent to which errors in performance lead to shifts in attention. We used two popular models of categorization to generate specific predictions about the extent to which attentional shifts would be biased to occur after error trials and not after correct trials. Simulations of the models using both random parameter values and best-fitting parameter values reveal consistent predictions: high error bias values are most likely, and a wide distribution of error biases are possible. We then examined error bias distributions from a variety of human eye-tracking studies of categorization. Across 8 data sets and 384 participants, distributions of error bias scores in humans were remarkable both in their consistency and in their contrast to model predictions. Human error biases were clustered around zero, and both high and low values were quite rare. The ease of implementing optimal attention seems to influence error biases, but otherwise, category structure does not influence scores. While error-driven accounts are clearly applicable in certain instances (Kopp & Wolff, 2000), the data presented here is a stark demonstration that under a variety of experimental circumstances mistakes do not have an immediate effect on the allocation of attention.

The present work is an attempt to understand how overt attentional allocation (i.e., gaze) changes over time and in response to classification feedback. There are currently no existing theories of category learning explicitly designed to account for attentional changes in category learning at the level of eye-movements. We used existing theories of attentional learning in categorization to guide our initial hypotheses about how such attentional learning might occur, while acknowledging that the modelers did not have eye-movements specifically in mind when they developed their theory. While there is some indication from previous studies that general trends in attentional learning are captured by these models, our data suggests that the algorithms used in extant theories of categorization are not well suited to changes in trial-by-trial attentional allocation.

Our measure, the error bias, captures the immediate effects of error. While our data are quite strong in suggesting that errors do not cause a lot more shifting of attention than correct trials, it is entirely possible, if not likely, that error still plays a role in the shifting of attention. The effects of error may accumulate, or be delayed in time. It is also possible that the learning of the categories is influenced by immediate errors, but that attentional allocation is a function of category knowledge and is thus only indirectly influenced by errors.

These data suggest that learning algorithms in extant models are not suitable for modeling eye-movements. Mixing supervised and unsupervised learning algorithms provides one possible solution. This approach has been taken within the LEABRA framework (O'Reilly, 1998) that incorporates Hebbian learning processes freeing the models from reliance upon an error signal. There has also been a recent interest in the various flavors of reinforcement learning, such as temporal difference learning (Holroyd & Coles, 2002; Jones & Cañas, 2010; Phillips & Noelle, 2004) or actor-critic models (Alexander, 2007), partially for reasons of biological plausibility. While these algorithms are error-driven, they accumulate a reinforcement history which may serve to diffuse the impact of trial-by-trial error on attentional shifting, spreading learning more evenly across trials. Our lab is currently investigating this possibility.

The problem of identifying the role of selective attention in learning has broad implications. One of the key benefits of selective attention is the reduction in complexity of an information source by biasing the selection of relevant information. Haider and Frensch (1999) argue that in a wide variety of tasks, performance is augmented when processing is limited to task-relevant properties. Research has also implicated impaired ability to selectively attend in a variety of clinical contexts. Greenaway and Plaisted (2005) suggest that autistic children's distractibility is related to their inability to selectively attend to stimuli with certain properties. In more traditional empirical work, Blair and Homa (2005) demonstrate how training with incomplete sources of information can lead to selective attention patterns which hinder performance relative to participants with no prior training at all. These studies highlight the importance of developing models that accurately capture the processes of selective attention. The work presented here makes progress toward that goal by providing strong evidence that popular conceptions of error-driven attentional learning are unsuitable for modeling eye-movements.

Acknowledgments

We thank the members of the Cognitive Science Lab for assistance with many aspects of this research. We also thank John Kruschke for generously supplying the RASHNL source code. This research was made possible by grants from the Natural Sciences and Engineering Research Council of Canada, the Canada Foundation for Innovation, and the British Columbia Knowledge Development Fund.

References

Alexander, W. H. (2007). Shifting attention using a temporal difference prediction error and high-dimensional input. *Adaptive Behavior*, 15(2), 121–133.

Ashby, F.G. & Gott, R.E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. Journal of Experimental *Psychology: Learning, Memory, and Cognition*, 14(1), 33-53.

Blair, M.R., Watson, M.R., Walshe, R.C., & Maj, F. (2009). Extremely selective attention: Eye-tracking studies on dynamic attentional allocation to stimulus features. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(5), 1196-1206. doi: 10.1037/a0016272

Blair, M.R., Watson, M.R., & Meier, K.M. (2009). Errors, efficiency, and the interplay between attention and category learning. *Cognition*, 112(2), 330-336. doi: 10.1016/j.cognition. 2009.04.008

Blair, M., & Homa, D. L. (2005). Integrating novel dimensions to eliminate category exceptions: When more is less. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(2), 258–271.

Biederman, I., & Shiffrar, M. M. (1987). Sexing day-old chicks: a case study and expert systems analysis of a difficult perceptual-learning task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 640–645.

Bott, L., Hoffman, A., & Murphy, G. L. (2007). Blocking in category learning. *Journal of Experimental Psychology: General*, 136, 685-699.

Greenaway R., Plaisted, K. (2005). Top-down attentional modulation in autistic spectrum disorders is stimulus-specific. *Psychological Science*, 16, 987-994.

Haider, H., & Frensch, P. A. (1999). Eye movement during skill acquisition: More evidence for the information-reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 172-190.

Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, 109(4), 679–709.

Jones, M., & Cañas, F. (2010). Integrating reinforcement learning with models of representation learning. *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*.

Kim, S., & Rehder, B. (2009). How prior knowledge affects selective attention during category learning: An eyetracking study. (Submitted for publication)

Kopp, B., & Wolff, M. (2000). Brain mechanisms of selective learning: Event-related potentials provide evidence for error-driven learning in humans. *Biological Psychology*, 51, 223–246.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, 99(1), 22–44.

Kruschke, J. K., & Johansen, M.K. (1999). A model of probabilistic category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(5), 1083-1119.

Little, D. R., & Lewandowsky, S. (2009). Beyond non-utilization: Irrelevant cues can gate learning in probabilistic categorization. *Journal of Experimental Psychology: Human Perception and Performance*, 35(2), 530–550.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, 111, 309-332.

Maddox, W.T. & Ashby, F.G. (2004). Dissociating explicit and procedural-learning base systems of perceptual category learning. *Behavioural Processes*, 66(3), 309-332.

Maddox, W.T., Filoteo, J.V., Hejl, K.D., & Ing, A.D. (2004). Category number impacts rule-based but not information-integration category learning: Further evidence for dissociable category-learning systems. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 30(1), 227-245.

McPeek, R.M., Maljkovic, V., & Nakayama, K. (1999). Saccades require focal attention and are facilitated by a short-term memory system. *Vision Research,* 39(8), 1555-1566.

O'Reilly, R. C. (1998). Six principles for biologically-based computational models of cortical cognition. *Trends in Cognitive Sciences*, 2(11), 455–462.

Phillips, J. L., & Noelle, D. C. (2004). Reinforcement learning of dimensional attention for categorization. In K. D. Forbus & D. Gentner & T. Regier (Eds.), *26th Annual Meeting of the Cognitive Science Society*, CogSci2004. Hillsdale, NJ: Erlbaum Publisher.

Quinn, P. C., Doran, M. M., Reiss, J. E., & Hoffman, J. E. (2009). Time course of visual attention in infant categorization of cats versus dogs: Evidence for a head bias as revealed through eye tracking. *Child Development*, 80, 151-161.

Rakison, D., & Butterworth, G. E. (1998). Infants' use of object parts in early categorization. *Developmental Psychology*, 34(1), 49-62.

Rehder, B., & Hoffman, A. B. (2005a). Eyetracking and selective attention in category learning. *Cognitive Psychology*, 51(1), 1-41. doi:10.1016/j.cogpsych. 2004.11.001

Rehder, B., & Hoffman, A. B. (2005b). Thirty-something category learning results explained: Selective attention, eyetracking, and models of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(5), 811-829. doi: 10.1037/0278-7393.31.5.811

Robert, S., & Pashler, H. (2000). How persuasive is a good fit? a comment on theory testing in psychology. *Psychological Review*, 107, 358-367.